

(19) World Intellectual Property Organization
International Bureau



(19)

(43) International Publication Date
19 April 2007 (19.04.2007)

PCT

(10) International Publication Number
WO 2007/044471 A2

- (51) International Patent Classification: Not classified
(21) International Application Number: PCT/US2006/038988
(22) International Filing Date: 5 October 2006 (05.10.2006)
(25) Filing Language: English
(26) Publication Language: English
(30) Priority Data: 11/244,826 6 October 2005 (06.10.2005) US
(71) Applicants (for all designated States except US): MASSACHUSETTS INSTITUTE OF TECHNOLOGY [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US). PARIVID LLC [US/US]; One Broadway, Suite 600, Cambridge, MA 02142 (US).
(72) Inventors; and
(75) Inventors/Applicants (for US only): BOSQUES, Carlos [US/US]; 71 Fulkerson Street, #106, Cambridge, MA 02141 (US). KEISER, Nishla [US/US]; 46 Lopez Street, Cambridge, MA 02139 (US). SRINIVASAN, Aravind [IN/US]; 99 Florence Street, #309, Malden, MA 02148 (US). RAMAN, Rahul [IN/US]; 6 Jason Street, #403, Arlington, MA 02476 (US). VISWANATHAN, Karthik [IN/US]; 474 Broadway, #16, Somerville, MA 02145 (US). SASISEKHARAN, Ram [US/US]; 4 Duval Way,

Bedford, MA 01730 (US). GANDHE, Pankaj [IN/US]; 91 Windingwood Drive, #6B, Sayreville, NJ 08872 (US). RAGURAM, Sasi [US/US]; 34 Vliet Drive, Hillsborough, NJ 08844 (US).

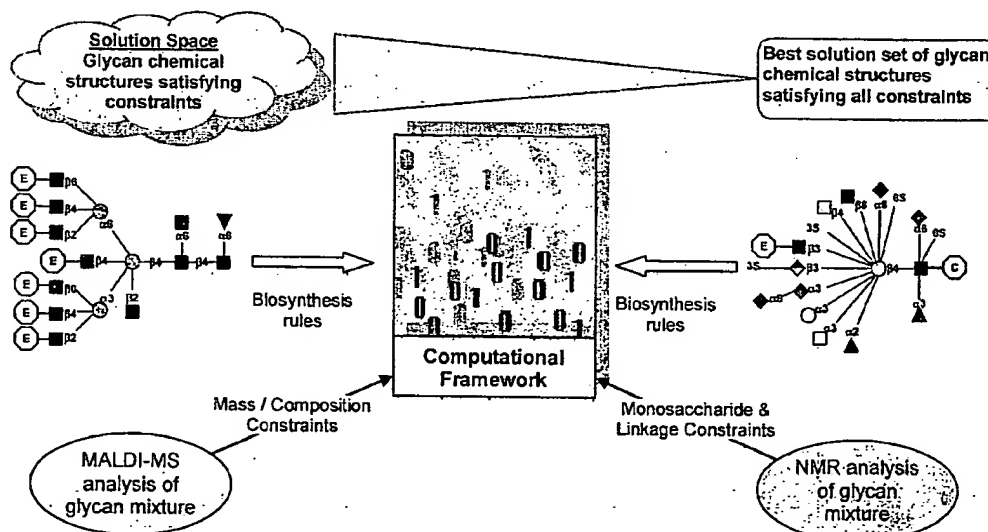
(74) Agent: VATLAND, Janice, A.; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: METHODS AND PRODUCTS RELATED TO THE IMPROVED ANALYSIS OF CARBOHYDRATES



(57) Abstract: The invention relates, in part, to the improved analysis of carbohydrates. In particular, the invention relates to the analysis of carbohydrates, such as N-glycans and O-glycans found on proteins and saccharides attached to lipids. Improved methods, therefore, for the study of glycosylation patterns on cells, tissue and body fluids are also provided. Information from the analysis of glycans, such as the glycosylation patterns on cells, tissues and in body fluids, can be used in diagnostic and treatment methods as well as for facilitating the study of the effects of glycosylation/alterd glycosylation. Such methods are also provided. Methods are further provided to assess production processes, to assess the purity of samples containing glycoconjugates, and to select glycoconjugates with the desired glycosylation.

WO 2007/044471 A2

PCT/US2007/044471 A2



Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS AND PRODUCTS RELATED TO THE IMPROVED ANALYSIS OF CARBOHYDRATES

5

FIELD OF THE INVENTION

The invention relates to the improved analysis of carbohydrates. In particular, the invention relates to the analysis of carbohydrates, such as N-glycans and O-glycans found on proteins and saccharides attached to lipids. The invention also relates to the analysis of glycoconjugates, such as glycoproteins, glycolipids and proteoglycans. Methods for the study of glycosylation patterns on cells, tissues and in body fluids, such as serum, are also provided. Information regarding the glycosylation patterns on cells, in tissues and in body fluids can be used in diagnostic and treatment methods as well as for facilitating the study of the effects of glycosylation/altered glycosylation on diseases, protein or lipid function as well as on the function of medical treatments. Information regarding the glycosylation of glycoconjugates can also be used in the quality control analysis of the production of glycoconjugates and therapeutics.

BACKGROUND OF THE INVENTION

Protein glycosylation, the attachment of carbohydrates to proteins, is one of the most common modifications found in eukaryotics. Glycosylation falls into three categories: N-linked modification of the asparagine (Asn) side chain, O-linked modification of serine (Ser) or threonine (Thr) and the modification of the protein C-carboxyl terminus by glycosylphosphatidyl inositol (GPI) derivatization. O-linked glycosylation and GPI anchor derivatization are post-translational modifications that take place in the Golgi. On the other hand, N-linked glycosylation is a co-translational modification. As proteins are synthesized, the polypeptide enters the endoplasmic reticulum, where oligosaccharyl transferase (OT) attaches a branched carbohydrate (N-glycan) to the side chain of certain asparagine residues (Hirschberg, C.B., Snider, M.D. 1987. *Annu Rev Biochem.* 56, 63-87.) This process requires an Asn-X-Ser/Thr consensus sequence in the peptide substrate, where X is any amino acid except proline (Bause, E. 1983. *Biochem J.* 209, 331-6; Marshall, R.D. 1972. *Annu Rev Biochem.* 41, 673-702.) The attached glycans, are subsequently modified by a complex array of glycosidases and glycosyl transferases in the endoplasmic reticulum (ER) and Golgi apparatus. The attached glycans play an important role in protein folding, as well as directing the protein to the appropriate location within the cell (Dwek, R.A. 1996. *Chem Rev.* 96, 683-

- 720; O'Connor, S.E., Imperiali, B. 1996. *Chem Biol.* 3, 803-12.) Outside the cell, the sugars aid in protein-protein interactions, often modulating the activity of the protein to which they are attached. Depending on the glycan composition, they can also protect against or facilitate protein degradation in circulation, as well as target the protein to a specific organ (Crocker, P.R., Varki, A. 2001. *Immunology.* 103, 137-45; Helenius, A., Aebi, M. 2001. *Science.* 291, 2364-9; Imperiali, B., O'Connor, S.E. 1999. *Curr Opin Chem Biol.* 3, 643-9.)

Glycans also have an important role in normal biology, as evidenced by the high lethality in cases of defective glycosylation. In mouse knockout models, disrupting even one of the biosynthetic enzymes can lead to enormous multisystemic disorders, and several result in embryonic lethality (Furukawa, K., et al. 2001. *Biochim Biophys Acta.* 1525, 1-12.) There are currently six recognized human congenital disorders of glycosylation (CDGs), all resulting in patients with multiple organ abnormalities, developmental delay and immune problems, among others (Jaeken, J., Matthijs, G. 2001. *Annu Rev Genomics Hum Genet.* 2, 129-51; Freeze, H.H., Aebi, M. 1999. *Biochim Biophys Acta.* 1455, 167-78; Carchon, H., et al. 1999. *Biochim Biophys Acta.* 1455, 155-65.) In fact, the immune system is one of the most commonly studied systems where glycans, such as N-glycans, have been shown to play an important physiological role. For example, specific carbohydrate structures are recognized by selectins, a family of proteins expressed on endothelial cells or lymphocytes that can trigger the immune system upon activation (Powell, L.D., et al. *J Biol Chem.* 268, 7019-27; Sgroi, D., et al. 1993. *J Biol Chem.* 268, 7011-8.) The same class of structures that are necessary for proper immune function can also provide a binding site for certain viruses, bacteria or tumor cells in the body (Karlsson, K.A. 1998. *Mol Microbiol.* 29, 1-11; Pritchett, T.J., et al. 1987. *Virology.* 160, 502-6.)

Viral infection is mediated by the interaction of viral proteins with glycans on the cell surfaces of the host (Van Eijk, M., et al. 2003. *Am J Respir Cell Mol Biol.* 6, 871-9.) Despite the increasing evidence associating glycans to different pathogenic conditions, in multiple instances it is unclear whether changes in glycan structure are a cause or a symptom of the disorder. In cystic fibrosis, increased antennary fucosylation (α 1-3 linked to GlcNAc) is observed on surface membrane glycoproteins of airway epithelial cells (Glick, M.C., et al. 2001. *Biochimie.* 83, 743-7; Scanlin, T.F., Glick, M.C. 2000. *Glycoconj. J* 17, 617-26.)

There have also been many reports of alterations in glycan composition on cancer cell proteins. For example, there are indications that prostate cancer cells produce prostate specific antigen (PSA) with more glycan branching than non-cancer cells (Peracaula, R., et al. 2003. *Glycobiology.* 13, 457-70; Belanger, A., et al. 1995. *Prostate.* 27, 187-97; Prakash,

S., Robbins, P.W. 2000. *Glycobiology* 10, 173-6.) Melanoma and bladder cancer cells produce proteins with highly branched glycans due to an overexpression of the biosynthetic enzyme β 1,6-*N*-acetyl-glucosaminyltransferase V (GnT-V) (Chakraborty, A.K., et al. 2001. *Cell Growth Differ.* 12, 623-30; Przybylo, M., et al. 2002. *Cancer Cell Int.* 2, 6.) Increased sialylation and additional branching have also been observed in cells from human breast and colon neoplasia (Lin, S., et al. 2002. *Exp Cell Res.* 276, 101-10; Nemoto-Sasaki, Y., et al. 2001. *Glycoconj J.* 18, 895-906; Dennis, J.W., et al. 1999. *Biochim Biophys Acta.* 1473, 21-34; Fernandes, B., et al. 1991. *Cancer Res.* 51, 718-23.)

SUMMARY OF THE INVENTION

This invention provides, in part, methods related to the analysis of carbohydrates. In particular, the invention relates to the analysis of carbohydrates, such as N-glycans and O-glycans found on proteins and saccharides attached to lipids. The invention also relates to the analysis of glycoconjugates, such as glycoproteins, glycolipids and proteoglycans.

In one aspect of the invention, therefore, a method of analyzing a sample containing one or more glycoconjugates, which comprise one or more carbohydrates (e.g., glycans) conjugated to a non-saccharide component is provided. The method, in one embodiment, includes the steps of analyzing the glycoconjugates to characterize the glycoconjugates, analyzing the non-saccharide components of the glycoconjugates to characterize the non-saccharide components, separating the carbohydrates (e.g., glycans) from the sample containing one or more glycoconjugates, analyzing the carbohydrates (e.g., glycans) to characterize the carbohydrates (e.g., glycans), and determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample with the results obtained from one or more of the analysis steps and a computational method. In another embodiment the computational method comprises generating constraints from the results obtained from one or more of the analysis steps and solving them.

In one embodiment the methods provided can include determining the glycosylation sites and glycosylation site occupancy of the glycoconjugates. In another embodiment the determination of the glycosylation sites and glycosylation site occupancy includes the steps of cleaving the non-saccharide components of the glycoconjugates, cleaving the carbohydrates (e.g., glycans) from the non-saccharide components and labeling the non-saccharide components of a first portion of the sample at the glycosylation sites, cleaving the carbohydrates (e.g., glycans) from the non-saccharide components of a second portion of the

sample, analyzing the first and second portions of the sample containing the non-saccharide components, and comparing the results.

In another embodiment the methods are also directed to matching one or more carbohydrates (e.g., glycans) (e.g., of a glycome) to a glycoconjugate. The method includes, in some embodiments, determining the glycosylation sites and glycosylation site occupancy of one or more glycoconjugates and determining the possible carbohydrates (e.g., glycans) at each site. The method can also include characterizing the glycome (i.e., characterizing the carbohydrates (e.g., glycans), glycoconjugates and/or components thereof). The method can also include the use of a computational method to match the carbohydrates (e.g., glycans) to the glycoconjugates. In one embodiment determining all possible carbohydrates (e.g., glycans) at each site includes comparing unlabeled to labeled glycoconjugates and unlabeled to labeled deglycosylated fragments of the glycoconjugates. In another embodiment the computational method includes generating constraints from the results of the characterization of the glycome and/or other information (i.e., one or more sets of data).

In one embodiment non-saccharide components of a first portion of a sample are labeled with a labeling agent. In another embodiment the labeling agent is an isotope of C, N, H, S or O. In still another embodiment the labeling agent is ^{18}O . In yet another embodiment the labeling agent is ^2H . In still another embodiment non-saccharide components of a second portion of a sample are unlabeled. In a further embodiment non-saccharide components of a second portion of a sample are labeled.

In one embodiment glycosylation site occupancy is quantified from ratios of the masses of the non-saccharide components of a first and second portion of a sample. In another embodiment a first and second portion of a sample are analyzed with a mass spectrometric method. In a further embodiment a first and second portion of a sample are analyzed separately. In yet another embodiment a first and second portion of a sample are analyzed as a mixture.

In another aspect of the invention a method of analyzing a sample containing one or more glycoconjugates, which comprise one or more carbohydrates (e.g., glycans) conjugated to a non-saccharide component, which includes analyzing the glycoconjugates to determine the glycosylation sites and glycosylation site occupancy, separating the carbohydrates (e.g., glycans) from the sample containing one or more glycoconjugates, analyzing the carbohydrates (e.g., glycans) to characterize the carbohydrates (e.g., glycans), and determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample is provided. In one embodiment determining the glycosylation sites and

glycosylation site occupancy comprises cleaving the carbohydrates (e.g., glycans) from the non-saccharide components and labeling the non-saccharide components at their glycosylation sites of a first portion of the sample, cleaving the carbohydrates (e.g., glycans) from the non-saccharide components of a second portion of the sample, analyzing the first and second portions of the sample of glycoconjugates and comparing the results. In another embodiment determining the glycosylation sites comprises analyzing the non-saccharide components to characterize the non-saccharide components. In still another embodiment determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample comprises generating constraints from the results of one or more of the analysis steps and solving the constraints.

In yet another aspect of the invention a method of analyzing a sample containing one or more carbohydrates is provided. In one embodiment the method includes analyzing the carbohydrates with MALDI-MS to determine the monomer composition and relative abundance of the carbohydrates, analyzing the carbohydrates with NMR to determine the monomer composition and linkage abundance of the carbohydrates, and generating constraints from the results of one or both of the analysis steps and solving the constraints with a computational method. In one embodiment the NMR is used to determine the relative abundance of one or more monomers or ratios of monomers. In still another embodiment the method further comprises analyzing the non-saccharide components of one or more glycoconjugates or a combination thereof, when the carbohydrates are part of one or more glycoconjugates.

In still another aspect of the invention a method of analyzing a sample containing carbohydrates is provided, which includes separating neutral from charged carbohydrates, and analyzing the neutral and charged carbohydrates separately to characterize the carbohydrates. In another embodiment the method, when the carbohydrates are part of one or more glycoconjugates, includes denaturing the glycoconjugates, separating the carbohydrates (e.g., glycans) from the non-saccharide components, and analyzing the carbohydrates (e.g., glycans).

In still another aspect of the invention a method of analyzing a sample containing one or more carbohydrates, which includes analyzing the carbohydrates in the presence of a thymine derivative and an ion exchange resin is provided.

The methods provided herein, in one embodiment, when the carbohydrates are part of glycoconjugates, can also include denaturing the glycoconjugates, separating the

carbohydrates (e.g., glycans) from the non-saccharide components or analyzing the carbohydrates (e.g., glycans) or some combination thereof.

In still another aspect of the invention a method of analyzing a sample containing carbohydrates (e.g., glycans) is provided, which includes analyzing a first portion of the sample, wherein the carbohydrates (e.g., glycans) have been removed, analyzing a second portion of the sample, wherein the second portion of the sample contains intact glycoconjugates, which comprise one or more carbohydrates (e.g., glycans) conjugated to a non-saccharide component, and analyzing a third portion of the sample, wherein the third portion of the sample contains carbohydrates (e.g., glycans).

In one embodiment the second portion of the sample containing intact glycoconjugates is analyzed with a method, which includes analyzing the glycoconjugates to characterize the glycoconjugates, analyzing the non-saccharide components of the glycoconjugates to characterize the non-saccharide components, separating the carbohydrates (e.g., glycans) from the sample containing one or more glycoconjugates, analyzing the carbohydrates (e.g., glycans) to characterize the carbohydrates (e.g., glycans), and determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample with the results obtained from one or more of the analysis steps and a computational method. In another embodiment the second portion of the sample containing intact glycoconjugates is analyzed with a method, which includes analyzing the glycoconjugates to determine the glycosylation sites and glycosylation site occupancy, separating the carbohydrates (e.g., glycans) from the sample containing one or more glycoconjugates, analyzing the carbohydrates (e.g., glycans) to characterize the carbohydrates (e.g., glycans), and determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample. In one embodiment determining the glycosylation sites and glycosylation site occupancy comprises cleaving the carbohydrates (e.g., glycans) from the non-saccharide components and labeling the non-saccharide components at their glycosylation sites of a first portion of the sample, cleaving the carbohydrates (e.g., glycans) from the non-saccharide components of a second portion of the sample, analyzing the first and second portions of the sample of glycoconjugates and comparing the results. In another embodiment the second portion of the sample containing intact glycoconjugates is analyzed with a method, which includes denaturing the glycoconjugates, separating the carbohydrates (e.g., glycans) from the non-saccharide components of the glycoconjugates, analyzing the carbohydrates (e.g., glycans) with MALDI-MS to determine the monomer composition and relative abundance of the carbohydrates (e.g., glycans), analyzing the carbohydrates (e.g., glycans) with NMR to

determine the monomer composition and linkage abundance of the carbohydrates (e.g., glycans), and generating constraints from the results of the analysis steps and solving the constraints with a computational method. In one embodiment the NMR is used to determine the relative abundance of one or more monomers or ratios of monomers. In still another embodiment the second portion of the sample containing intact glycoconjugates is analyzed with a method, which comprises denaturing the glycoconjugates, separating the carbohydrates (e.g., glycans) from the non-saccharide components of the glycoconjugates, separating neutral from charged carbohydrates (e.g., glycans), and analyzing the neutral and charged carbohydrates (e.g., glycans) separately to characterize the carbohydrates (e.g., glycans). In yet another embodiment the second portion of the sample containing intact glycoconjugates is analyzed with a method, which comprises denaturing the glycoconjugates, separating the carbohydrates (e.g., glycans) from the non-saccharide components of the glycoconjugates, and analyzing the carbohydrates (e.g., glycans) in the presence of a thymine derivative and an ion exchange resin.

In yet a further embodiment the third portion of the sample containing carbohydrates (e.g., glycans) is analyzed with a method, which includes analyzing the carbohydrates (e.g., glycans) with MALDI-MS to determine the monomer composition and relative abundance of the carbohydrates (e.g., glycans), analyzing the carbohydrates (e.g., glycans) with NMR to determine the monomer composition and linkage abundance of the carbohydrates (e.g., glycans), and generating constraints from the results of one or more of the analysis steps and solving the constraints with a computational method. In one embodiment the NMR is used to determine the relative abundance of one or more monomers or ratios of monomers. In yet a further embodiment the third portion of the sample containing carbohydrates (e.g., glycans) is analyzed with a method, which comprises separating neutral from charged carbohydrates (e.g., glycans), and analyzing the neutral and charged carbohydrates (e.g., glycans) separately to characterize the carbohydrates (e.g., glycans). In yet a further embodiment the third portion of the sample containing carbohydrates (e.g., glycans) is analyzed with a method, which comprises analyzing the carbohydrates (e.g., glycans) in the presence of a thymine derivative and an ion exchange resin. In still another embodiment the carbohydrates (e.g., glycans) of the third portion of the sample are not part of intact glycoconjugates. In another embodiment the carbohydrates (e.g., glycans) of the third portion of the sample are part of intact glycoconjugates, which comprise one or more carbohydrates (e.g., glycans) conjugated to a non-saccharide component, and the method further includes denaturing the

glycoconjugates, and separating the carbohydrates (e.g., glycans) from the non-saccharide components of the glycoconjugates.

In another embodiment the methods provided can include determining the sequence of one or more non-saccharide components. In one embodiment the non-saccharide components are peptides and a peptide sequence is determined. In yet another embodiment the sequence of one or more non-saccharide components is determined prior to or subsequent to analysis of one or more glycoconjugates.

The methods provided can also include the generation of constraints. Constraints can be generated with results from an analysis of a carbohydrate, glycoconjugate or a component thereof (i.e., glycan or non-saccharide component) with an analytical (or experimental) method and/or by using other information. In one embodiment constraints are generated from databases containing information about carbohydrates (e.g., glycans), non-saccharide components, glycoconjugates or a combination thereof. In another embodiment constraints are generated from biosynthetic rules. In still another embodiment constraints are generated from information about the sample origin. In one embodiment the information about the sample origin comprises information regarding the expression system or expression conditions for the synthesis of carbohydrates, glycoconjugates or components thereof, the species from which carbohydrates, glycoconjugates or components thereof are derived, the expression levels of glycosidases and glycosyltransferases from the source from which carbohydrates, glycoconjugates or components thereof are obtained, or the state of the source from which the carbohydrates, glycoconjugates or components thereof are obtained. In another embodiment constraints can also be generated from the results of another experimental method. In one embodiment the other experimental method is a mass spectrometric method, an electrophoretic method, a NMR method, a chromatographic method or some combination thereof. In another embodiment the other experimental method is different from the first experimental method. In one embodiment, when the first experimental method is MALDI-MS, the other experimental method is not MALDI-MS. In another embodiment, where the first experimental method is NMR, the other experimental method is a different NMR method.

In one embodiment the constraints are one or more logic expressions, such as one or more mathematical equations. Preferably, in another embodiment, the constraints are more than one mathematical equation. Even more preferably, in still another embodiment, the constraints are more than two mathematical equations. In one embodiment, therefore, the constraints are three or more mathematical equations. In another embodiment the constraints

are five or more mathematical equations. In another embodiment the constraints are solved by determining the solution of the one or more mathematical equations. In still another embodiment the constraints are solved with a computer program.

The samples or portions thereof and carbohydrates, glycoconjugates or components thereof can be analyzed using any of a number or combination of experimental methods. In one embodiment two or more experimental methods can be used for analysis. In one embodiment the experimental method is a mass spectrometric method, an electrophoretic method, NMR, a chromatographic method or a combination thereof. In another embodiment the mass spectrometric method is LC-MS, LC-MS/-MS, MALDI-MS, MALDI-TOF-MS, MALDI-TOF PSD-MS, MALDI-TOF/TOF-MS, MALDI-TOF/TOF-MS/MS, MALDI-TOF/TOF PSD-MS, MALDI-FTMS, LC-MALDI-TOF/TOF-MS, Nano-LC MALDI-TOF/TOF-MS, Nano-LC MALDI-TOF/TOF PSD-MS, Nano-LC MALDI-TOF/TOF-MS/MS or TANDEM-MS. In yet another embodiment the mass spectrometric method is ESI-MS, LC-MS, LC-MS/-MS, MALDI-MS, MALDI-MS/MS, MALDI-TOF-MS, MALDI-TOF PSD-MS, MALDI-TOF/TOF-MS, MALDI-TOF/TOF-MS/MS, MALDI-TOF/TOF PSD-MS, MALDI-FTMS, LC-MALDI-TOF/TOF-MS, Nano-LC MALDI-TOF/TOF-MS, Nano-LC MALDI-TOF/TOF PSD-MS, Nano-LC MALDI-TOF/TOF-MS/MS or TANDEM-MS. In still another embodiment the mass spectrometric method is LC-MS, LC-MS/-MS, LC-FTMS, TANDEM-MS, MALDI-MS, MALDI-TOF-TOF-MS, MALDI-FTMS or MALDI/PSD-MS. In yet another embodiment the mass spectrometric method is a quantitative MALDI-MS, MALDI-TOF-TOF-MS or MALDI-FTMS using optimized conditions. In still another embodiment the experimental method is MALDI-MS. In one embodiment the MALDI-MS provides monomer (e.g., monosaccharide) composition and relative abundance information. In another embodiment MALDI-MS is used with another experimental method.

In still another embodiment the experimental method is nuclear magnetic resonance (NMR). In one embodiment the results from the NMR provide monomer (e.g., monosaccharide) composition and linkage information. In one embodiment the NMR is used to determine the relative abundance of one or more monomers or ratios of monomers.

In another embodiment the experimental method is NMR or MALDI-MS. In still a further embodiment both NMR and MALDI-MS is used for analysis.

In another embodiment the electrophoretic method is capillary electrophoresis (CE) or CE-LIF. In yet another embodiment the chromatographic method is HPLC.

The analysis in the methods provided can include the use of a mass spectrometric method, such as MALDI-MS, in the presence of a thymine derivative and an ion exchange

resin. In one embodiment the thymine derivative is thiothymine, 2-thiothymine, 4-thiothymine, 5-aza-2-thiothymine or 6-aza-2-thiothymine (ATT). In another embodiment the ion exchange resin is an ammonium resin, a cationic exchange resin, a cationic exchange resin in pyridinium form, an anionic exchange resin or a perfluorinated ion exchange resin.

5 In still another embodiment the perfluorinated ion exchange resin is NafionTM.

The methods provided can also include contacting the carbohydrates with one or more carbohydrate-degrading enzymes. In one embodiment the one or more carbohydrate-degrading enzymes are glycan-degrading enzymes, which include, for example, sialidase, galactosidase, mannosidase, N-acetylhexosaminidase or a combination thereof. In another
10 embodiment the methods provided can also include contacting the carbohydrates with strong acidic or basic conditions.

The methods provided can also include quantifying the carbohydrates using calibration curves of known carbohydrate standards.

The methods provided can also include purification steps. In one embodiment the
15 carbohydrates (e.g., glycans) are purified with solid phase extraction cartridges or ion exchange resins. In another embodiment the solid phase extraction cartridges are graphitic carbon columns, non-graphitic carbon columns or C-18 columns.

The methods provided can also include cleavage steps, which include the use of PNGase F, Endo H, Endo F, hydrazinolysis or alkaline borohydride.

20 The methods provided can also include denaturation with a denaturing agent. In one embodiment the denaturing agent comprises a detergent, urea, high salt concentration, guanidium hydrochloride or heat.

In another embodiment the methods provided can also include reduction with a reducing agent following denaturation. In one embodiment the reducing agent comprises
25 dithiothreitol (DTT), β -mercaptoethanol or Tris(2-carboxyethyl)phosphine (TCEP).

In still another embodiment the methods provided can include alkylation with an alkylating agent following reduction. In one embodiment the alkylating agent is iodoacetic acid or iodoacetamide.

The carbohydrates analyzed by the methods provided herein can be any carbohydrate
30 or combination of carbohydrates. In one embodiment the carbohydrates are polysaccharides. In still another embodiment the carbohydrates are glycans. In a further embodiment the carbohydrate is a glycosaminoglycan. In yet another embodiment the carbohydrate is hyaluronic acid. In another embodiment the carbohydrates are branched. In yet another embodiment they are unbranched. In still another embodiment the carbohydrates are a

mixture of branched and unbranched carbohydrates. In a further embodiment the carbohydrates are a mixture of a number of different carbohydrates. In another embodiment the carbohydrates are conjugated to a non-saccharide component and form one or more glycoconjugates. In one embodiment the glycoconjugate is a peptide-based glycoconjugate.

- 5 In another embodiment the glycoconjugate is a lipid-based glycoconjugate. In another embodiment the carbohydrates are modified or unmodified or a mixture thereof. In one embodiment the carbohydrates are glycans that are modified by permethylation or conjugation to a peptide.

- The methods provided can also include generating a list of all possible carbohydrates
10 (e.g., glycans) and/or glycoforms. In one embodiment the list is based on the results of the analysis of the carbohydrates (e.g., glycans), glycoconjugates, database information, biosynthetic rules information or a combination thereof.

- The methods provided can also include removing abundant or nonglycosylated lipids and/or proteins from a sample. In one embodiment the abundant or nonglycosylated lipids or
15 proteins are albumins or immunoglobulins.

The methods provided in one embodiment have a detection limit of less than about 1000, 500, 100, 75, 50, 25, 20, 18, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 femtomole(s).

- The methods provided in another embodiment can be used to detect low abundance
20 species. In one embodiment the low abundance species include, but are not limited to, glycans that contain fucoses, sialic acids, galactoses, mannoses or sulfate groups.

- The methods provided can be performed in a high-throughput manner. Therefore, in one embodiment a method or a portion thereof is performed in a 96-well plate or with a protein-binding membrane. In one embodiment the 96-well plate comprises a protein-
25 binding membrane. In another embodiment the protein-binding membrane is a polyvinylidene difluoride (PVDF) membrane, C-18 membrane or a nitrocellulose membrane.

- The methods provided can also be performed on a sample (or more than one sample) of carbohydrates, glycoconjugates or components thereof that are in solution or are immobilized on a solid support. In one embodiment the solid support is in a 96-well plate
30 format or comprises a membrane. In another embodiment the membrane is a protein-binding membrane. In one embodiment the membrane is a polyvinylidene difluoride (PVDF) membrane, C-18 membrane or a nitrocellulose membrane.

The methods provided can include the use of robotics. In one embodiment one or more steps of the methods provided or one or more portions thereof are performed with the use of robotics.

5 In another embodiment neutral and charged carbohydrates are analyzed separately in the methods provided.

The methods provided can be performed on any sample containing one or more carbohydrates. In one embodiment the sample is a sample comprising one or more glycoconjugates, one or more cells, a tissue or body fluid from a subject. In another embodiment the sample is a sample of serum, plasma, blood, urine, saliva, sputum, tears, 10 cerebral spinal fluid (CSF), seminal fluid, feces, tissues or cells. In still another embodiment the sample of body fluid is from a subject with a disease or condition. In yet another embodiment the sample of body fluid is from a subject that is undergoing treatment for a disease. In still another embodiment the sample of body fluid is from a healthy subject. In another embodiment the sample of body fluid is from a pregnant woman. In another 15 embodiment the sample is a sample comprising one or more glycoconjugates. In still another embodiment the sample is a batch of glycoconjugates. In yet another embodiment more than one sample is analyzed. In one embodiment the more than one sample are two or more batches of glycoconjugates. In another embodiment the more than one sample are two or more samples containing carbohydrates (e.g., glycans). In still another embodiment the 20 sample(s) are contained in a 96-well plate or on a protein-binding membrane. In a further embodiment the sample(s) are in solution. In another embodiment the samples are analyzed as a mixture.

The methods provided can be used to analyze an entire sample or one or more portions or fractions thereof. In one embodiment an entire sample is analyzed. In another 25 embodiment a fraction of a sample is analyzed.

Therefore, the methods provided can also include the fractionation of a sample or portion thereof. In one embodiment the fractionation is based on charge, size, molecular weight, binding properties, acidity, basicity, pI, hydrophobicity or hydrophilicity. In another embodiment the fractionation is performed using solid supports with immobilized proteins, 30 organic molecules, inorganic molecules, lipids, carbohydrates or nucleic acids; filters or resins. In another embodiment carbohydrates of a fractionated part of a sample are the carbohydrates that are analyzed. In still another embodiment a fractionated part of a sample is isolated. In yet another embodiment a fraction of a sample is removed and it is the remaining fraction that is analyzed. In one embodiment the fraction of a sample removed

contains acidic carbohydrates (e.g., acidic glycans). In another embodiment the fraction of a sample removed contains neutral carbohydrates (e.g., neutral glycans). In still another embodiment the fraction of a sample removed contains high abundance proteins. In one embodiment the high abundance proteins are albumins or immunoglobulins. In another embodiment the high abundance proteins are immunoglobulins. In still another embodiment the fraction of a sample removed does not contain high abundance proteins.

Any of the methods provided herein can be used as a method of analyzing the purity of a sample.

Any of the methods provided herein can be used as a method for diagnosis or assessing prognosis.

Any of the methods provided herein can be used as a method of assessing the effectiveness of a treatment of a subject.

In still another aspect of the invention a method of generating a glycoconjugate library, wherein the glycoconjugate comprises one or more carbohydrates (e.g., glycans) conjugated to a non-saccharide component, is provided. In one embodiment the method includes cleaving the non-saccharide components of one or more glycoconjugates in a sample and labeling the non-saccharide component fragments generated with a labeling agent in order to generate a glycoconjugate library, and cleaving the carbohydrates (e.g., glycans) from the non-saccharide components and labeling the non-saccharide components in the sample at the glycosylation sites. In one embodiment the labeling agent is an isotope of C, N, H, S or O. In still another embodiment the labeling agent is ^{18}O . In still a further embodiment the labeling agent is ^2H . In yet another embodiment the method further comprises analyzing the fragments generated from the cleavage of the glycoconjugates.

In one embodiment the analysis is performed on any sample containing one or more glycoconjugates. In another embodiment the glycoconjugate is a lipid-based glycoconjugate. In yet another embodiment the glycoconjugate is a peptide-based glycoconjugate. In still another embodiment the analyzing results in the characterization of the glycosylation sites, the peptides and the carbohydrates (e.g., glycans) of the peptide-based glycoconjugate.

In yet another aspect of the invention a library generated with the methods provided herein is also provided.

In still another aspect of the invention a method of analyzing a sample containing glycoconjugates, is provided, which includes modifying the glycoconjugates in the sample, and comparing the modified glycoconjugates with the library. In one embodiment the glycoconjugates are modified by cleavage, labeling or both. In another embodiment the step

of modifying the glycoconjugates comprises cleaving the glycoconjugates to generate glycoconjugate fragments. In another embodiment the glycoconjugates are cleaved by cleaving the non-saccharide components of the glycoconjugates, cleaving the carbohydrates (e.g., glycans) of the glycoconjugates, cleaving the carbohydrates (e.g., glycans) from the non-saccharide components of the glycoconjugates or a combination thereof to generate fragments. In yet another embodiment the fragments generated are labeled. In one embodiment the non-saccharide component fragments are labeled. In a further embodiment the non-saccharide component fragments are labeled at the sites of cleavage. In another embodiment the non-saccharide component fragments are labeled at their glycosylation sites. In still another embodiment the step of comparing includes mixing the glycoconjugate fragments with the library and determining the ratios of the glycoconjugate fragments to the library. In one embodiment known proportions of samples of the glycoconjugate fragments are mixed with the library.

In yet a further aspect of the invention a method of generating a list of properties is provided. In one embodiment the method includes determining one or more properties of a sample with a method as provided herein, and recording a value for the one or more properties to generate a list, wherein the value of the one or more properties is recorded in a computer-generated data structure. In one embodiment the one or more properties comprise the number of one or more types of monomers of a carbohydrate in the sample. In another embodiment the one or more properties comprise the mass of a carbohydrate or portion thereof in the sample. In still another embodiment the one or more properties comprises the quantity of a carbohydrate or portion thereof in the sample. In one embodiment the carbohydrate is conjugated to a glycoconjugate. In another embodiment the glycoconjugate is a peptide-based glycoconjugate, and the one or more properties comprises the mass of the peptide of the peptide-based glycoconjugate. In another embodiment the glycoconjugate is a lipid-based glycoconjugate, and the one or more properties comprises the mass of the lipid of the lipid-based glycoconjugate. In still another embodiment the one or more properties comprises the mass of the glycoconjugate.

Also provided in another aspect of the invention is a database, tangibly embodied in a computer-readable medium, for storing information descriptive of one or more carbohydrates, the database, which includes one or more data units corresponding to the one or more carbohydrates, each of the data units including an identifier that includes one or more fields, each field for storing a value corresponding to one or more properties of the carbohydrates,

wherein the value corresponding to one or more properties of the carbohydrates is determined with a method as provided herein. In one embodiment the method includes analyzing the carbohydrates with MALDI-MS to determine the monomer composition and relative abundance of the carbohydrates, and analyzing the carbohydrates with NMR to determine the monomer composition and linkage abundance of the carbohydrates. In one embodiment NMR is used to determine the relative abundance of one or more monomers or ratios of monomers. In another embodiment the method also includes generating constraints from the results of one or more of the analysis steps and solving the constraints with a computational method. In yet another embodiment the methods include separating neutral from charged carbohydrates, and analyzing the neutral and charged carbohydrates separately to characterize the carbohydrates. In still another embodiment the method includes analyzing the carbohydrates in the presence of a thymine derivative and an ion exchange resin.

In one embodiment where the carbohydrates are part of intact glycoconjugates, which comprise one or more carbohydrates (e.g., glycans) conjugated to a non-saccharide component, the method further includes denaturing the glycoconjugates, and separating the carbohydrates (e.g., glycans) from the non-saccharide components of the glycoconjugates.

In another aspect of the invention a database, tangibly embodied in a computer-readable medium, for storing information descriptive of one or more glycoconjugates, the database, which includes one or more data units corresponding to the one or more glycoconjugates, each of the data units including an identifier that includes one or more fields, each field for storing a value corresponding to one or more properties of the glycoconjugates, wherein the value corresponding to one or more properties of the glycoconjugates is determined with a method as provided herein is provided. In one embodiment the method includes analyzing the glycoconjugates to characterize the glycoconjugates, analyzing the non-saccharide components of the glycoconjugates to characterize the non-saccharide components, separating the carbohydrates (e.g., glycans) from the sample containing one or more glycoconjugates, and analyzing the carbohydrates (e.g., glycans) to characterize the carbohydrates (e.g., glycans). In another embodiment the method also includes determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample with the results obtained from one or more analysis steps and a computational method. In another embodiment the method includes analyzing the glycoconjugates to determine the glycosylation sites and glycosylation site occupancy, separating the carbohydrates (e.g., glycans) from the sample containing one or more glycoconjugates, and analyzing the carbohydrates (e.g., glycans) to characterize the

carbohydrates (e.g., glycans). In one embodiment determining the glycosylation sites and glycosylation site occupancy comprises cleaving the carbohydrates (e.g., glycans) from the non-saccharide components and labeling the non-saccharide components at their glycosylation sites of a first portion of the sample, cleaving the carbohydrates (e.g., glycans) from the non-saccharide components of a second portion of the sample, analyzing the first and second portions of the sample of glycoconjugates and comparing the results. In still another embodiment the method also includes determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample. In yet another embodiment the method includes denaturing the glycoconjugates, separating the carbohydrates (e.g., glycans) from the non-saccharide components of the glycoconjugates, analyzing the carbohydrates (e.g., glycans) with MALDI-MS to determine the monomer composition and relative abundance of the carbohydrates (e.g., glycans), and analyzing the carbohydrates (e.g., glycans) with NMR to determine the monomer composition and linkage abundance of the carbohydrates (e.g., glycans). In one embodiment the NMR is used to determine the relative abundance of one or more monomers or ratios of monomers. In another embodiment the method also includes generating constraints from the results of one or more of the analysis steps and solving the constraints with a computational method. In still another embodiment the method includes denaturing the glycoconjugates, separating the carbohydrates (e.g., glycans) from the non-saccharide components of the glycoconjugates, separating neutral from charged carbohydrates (e.g., glycans), and analyzing the neutral and charged carbohydrates (e.g., glycans) separately to characterize the carbohydrates (e.g., glycans). In a further embodiment the method includes denaturing the glycoconjugates, separating the carbohydrates (e.g., glycans) from the non-saccharide components of the glycoconjugates, and analyzing the carbohydrates (e.g., glycans) in the presence of a thymine derivative and an ion exchange resin.

In another aspect of the invention a method of analyzing the total glycome of a sample is provided. In one embodiment the method includes analyzing all of the carbohydrates (e.g., glycans) of the sample, and determining a profile of the carbohydrates (e.g., glycans) of the sample. In another embodiment the composition of the carbohydrates (e.g., glycans) in the sample is determined. In still another embodiment the structures of the carbohydrates (e.g., glycans) in the sample are determined. In another embodiment the analysis of all of the carbohydrates (e.g., glycans) includes quantifying the carbohydrates (e.g., glycans) using calibration curves based on known carbohydrate (e.g., glycan) standards.

In another embodiment the profile of the carbohydrates (e.g., glycans) is a spectrum of monomer composition and relative abundance of the carbohydrates (e.g., glycans).

In one embodiment the carbohydrates (e.g., glycans) of the sample are analyzed with a mass spectrometric method, an electrophoretic method, NMR, a chromatographic method or a combination thereof. In another embodiment the analysis is performed, for example, with ESI-MS, LC-MS, LC-MS/MS, MALDI-TOF-MS, MALDI-MS/MS, MALDI-FTMS, TANDEM-MS, NMR, HPLC or CE. In a further embodiment the analysis is performed with MALDI-MS or MALDI-FTMS. In yet another embodiment the analysis is performed with electrophoresis, microfluidic devices or nanofluidic devices. In still another embodiment the carbohydrates (e.g., glycans) are further analyzed with another experimental method. In one embodiment the other experimental method provides linkage information. In another embodiment the other experimental method is LC-MS, LC-MS/MS, CE-LIF or NMR. In still another embodiment the other experimental method comprises the use of one or more carbohydrate-degrading enzymes (e.g., glycan-degrading enzymes). In still another embodiment the carbohydrates (e.g., glycans) of the sample are analyzed with a method, which includes analyzing the carbohydrates (e.g., glycans) with MALDI-MS to determine the monomer composition and relative abundance of the carbohydrates (e.g., glycans), and analyzing the carbohydrates (e.g., glycans) with NMR to determine the monomer composition and linkage abundance of the carbohydrates (e.g., glycans). In one embodiment the NMR is used to determine the relative abundance of one or more monomers or ratios of monomers. In another embodiment the method also includes generating constraints from the results of the analysis with MALDI-MS and NMR and solving the constraints with a computational method. In a further embodiment the carbohydrates (e.g., glycans) of the sample are analyzed with a method, which includes separating neutral from charged carbohydrates (e.g., glycans), and analyzing the neutral and charged carbohydrates (e.g., glycans) separately to characterize the carbohydrates (e.g., glycans). In still another embodiment the carbohydrates (e.g., glycans) of the sample are analyzed with a method, which includes analyzing the carbohydrates (e.g., glycans) in the presence of a thymine derivative and an ion exchange resin.

In one embodiment the carbohydrates are part of intact glycoconjugates, which comprise one or more carbohydrates (e.g., glycans) conjugated to non-saccharide components. In another embodiment the method also includes separating the carbohydrates (e.g., glycans) from the non-saccharide components of the glycoconjugates. In one embodiment the carbohydrates (e.g., glycans) are separated by cleavage with an enzymatic

method or a chemical method. In one embodiment the enzymatic method includes the use of PNGase F, Endo H or Endo F. In another embodiment the chemical method includes hydrazinolysis or the use of alkaline borohydride. In still another embodiment the cleavage is performed in a 96-well plate (e.g., with the glycoconjugates immobilized on a membrane), on
5 a protein-binding membrane or in solution. In still a further embodiment the cleavage is performed with the use of robotics or manually. In another embodiment the method further comprises purification of the carbohydrates (e.g., glycans). In one embodiment the purification is performed in a 96-well plate. In another embodiment the purification is performed using individual purification columns or cartridges. In yet another embodiment
10 the purification is performed with solid phase extraction cartridges. In still another embodiment the purification is performed with the use of robotics or manually.

In yet another embodiment the method includes analyzing the glycoconjugates to characterize the glycoconjugates, analyzing the non-saccharide components of the glycoconjugates to characterize the non-saccharide components, separating the carbohydrates
15 (e.g., glycans) from the sample, and analyzing the carbohydrates (e.g., glycans) to characterize the carbohydrates (e.g., glycans). In still another embodiment the method also includes determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample with the results obtained from the analysis and a computational method. In still another embodiment the method includes analyzing the glycoconjugates to determine the
20 glycosylation sites and glycosylation site occupancy, separating the carbohydrates (e.g., glycans) from the sample, and analyzing the carbohydrates (e.g., glycans) to characterize the carbohydrates (e.g., glycans). In one embodiment determining the glycosylation sites and glycosylation site occupancy comprises cleaving the carbohydrates (e.g., glycans) from the non-saccharide components and labeling the non-saccharide components at their
25 glycosylation sites of a first portion of the sample, cleaving the carbohydrates (e.g., glycans) from the non-saccharide components of a second portion of the sample, analyzing the first and second portions of the sample and comparing the results. In another embodiment the method also includes determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample.

30 In still another embodiment the method also includes identifying a pattern by performing a pattern analysis on the results using a computational method. In one embodiment the computational method is an iterative computational method. In another embodiment the iterative computational method determines the glycoforms in the sample. In still a further embodiment the method also includes recording one or more values

representing the pattern in a computer-generated data structure. In still another embodiment the method also includes associating the pattern with one or more samples of known origin (e.g., a sample from a diseased patient, a sample for one or more persons with one or more specific characteristics, a sample of a batch of glycoconjugates, etc.) In one embodiment, therefore, the pattern is associated with a population (e.g., healthy subjects, subjects with a specific disease, pregnant women, subjects that have specific demographic characteristics, etc.) In another embodiment the pattern is associated with a disease. In still another embodiment the pattern is associated with patterns of one or more samples of known origin. In one embodiment the pattern is associated by comparing the pattern with one or more patterns of one or more samples of known origin. In another embodiment the pattern is associated by extracting features of the pattern and comparing the features with information available for the one or more samples of known origin.

In one embodiment the pattern provides diagnostic or prognostic information. In another embodiment the pattern provides information about a sample, a person or population from which the sample was derived or sample origin.

In one embodiment the sample is from a subject and the pattern provides information about the subject's state. In another embodiment the subject's state is a diseased state.

In another embodiment the identified pattern is compared to the pattern of at least one other sample. In one embodiment the at least one other sample is a batch of glycoconjugates. In another embodiment the at least one other sample is a sample from a healthy or diseased individual.

In one embodiment the identified pattern is compared to another pattern. In another embodiment the other pattern is a known pattern. In still another embodiment the other pattern is an unknown pattern. In yet another embodiment the other pattern is a pattern that represents a batch of glycoconjugates. In one embodiment the method is a method to assess the purity of a batch of glycoconjugates and the known pattern represents a batch of glycoconjugates of known purity.

In still another embodiment the other pattern is a pattern that represents a diseased or healthy state. In one embodiment the diseased state is associated with cancer. In another embodiment the cancer is prostate cancer, melanoma, bladder cancer, breast cancer, lymphoma, ovarian cancer, lung cancer, colorectal cancer or head and neck cancer. In another embodiment the diseased state is associated with an immunological disorder. In still another embodiment the diseased state is associated with a neurodegenerative disease. In one embodiment the neurodegenerative disease is a transmissible spongiform encephalopathy,

Alzheimer's disease or neuropathy. In another embodiment the diseased state is associated with inflammation. In still another embodiment the diseased state is associated with rheumatoid arthritis. In yet another embodiment the diseased state is associated with cystic fibrosis. In a further embodiment the diseased state is associated with an infection. In one
5 embodiment the infection is viral or bacterial. In another embodiment the diseased state is associated with a congenital disorder.

In another embodiment the method is a method of monitoring prognosis and the known pattern is associated with the prognosis of a disease.

In still another embodiment the method is a method of monitoring drug treatment and
10 the known pattern is associated with a drug treatment.

In another embodiment the method also includes validating the association of the pattern. In one embodiment the association of the pattern is validated with one or more patterns of one or more samples of known origin. In another embodiment the pattern is validated by comparing it with one or more patterns of one or more samples of known origin.

15 In another aspect of the invention a method of generating a glycoprofile with a method provided herein is also provided.

In still another aspect of the invention a method of creating a database of glycoprofiles, which includes generating a glycoprofile of a sample according to a method provided, and recording one or more values corresponding to the glycoprofile in a computer-
20 generated data structure is provided. In yet another aspect of the invention the database so created is also provided.

In another aspect of the invention a method of determining a glycome pattern, which includes obtaining a glycoprofile of total carbohydrates (e.g., glycans) of a sample with a method provided herein, identifying features of the glycoprofile, generating data sets based
25 on the features of the glycoprofile, identifying a pattern in the data sets, and determining whether or not the pattern is associated with a known sample or diseased state is provided. In one embodiment the sample is obtained from a subject. In another embodiment the subject has a disease or condition. In another embodiment determining the glycoprofile includes obtaining more than one glycoprofile spectra. In one embodiment one of the spectra is of
30 acidic carbohydrates (e.g., acidic glycans). In another embodiment one of the spectra is of neutral carbohydrates (e.g., neutral glycans). In still another embodiment one spectra is of acidic carbohydrates and another spectra is of neutral carbohydrates.

In yet another embodiment when the analysis includes the generation of one or more glycoprofile spectra, the methods provided can also include assigning all of the possible carbohydrates (e.g., glycans) to the peaks of the one or more spectra.

In one embodiment the feature identified is the presence of one or more carbohydrates (e.g., glycans), the absence of one or more carbohydrates (e.g., glycans), the relative amount of one or more carbohydrates (e.g., glycans), the combination of two or more classes of carbohydrates (e.g., glycans), the presence of a specific carbohydrate (e.g., glycan) motif (i.e., a specific set of one or more monomers (e.g., monosaccharides)), the absence of a specific carbohydrate (e.g., glycan) motif, the relative amount of a specific carbohydrate (e.g., glycan) motif, the presence of one or more monomers in a carbohydrate (e.g., glycan), the absence of one or more monomers in a carbohydrate (e.g., glycan), the relative amount of one or more monomers in a carbohydrate (e.g., glycan) or the bond between monomers of a carbohydrate. In some embodiments combinations of features are identified.

In one embodiment the pattern is identified by linear discriminant, nearest neighbor, statistical classifier, neural net, decision tree, decision rules or association rules analysis.

In another embodiment the glycoprofile is generated from determining the glycosylation site occupancy of glycoconjugates in a sample. In still another embodiment the glycoprofile is determined by identifying and quantifying all of the carbohydrates (e.g., glycans) and/or glycoconjugates of the sample. In still another embodiment all of the carbohydrates (e.g., glycans) are identified and quantified by solving constraints with a computational method.

In another aspect of the invention a method of determining a glycome pattern of a sample, which includes determining the glycoprofile of the sample according to a method as provided herein, extracting one or more features of the glycoprofile, analyzing the one or more features, and validating the glycome pattern is provided. In one embodiment the one or more features is the presence or absence of a specific carbohydrate (e.g., glycan), an amount of a specific carbohydrate (e.g., glycan), a combination of specific carbohydrates (e.g., glycans), etc. In another embodiment the one or more features is a ratio between two or more carbohydrates (e.g., glycans) or monomers or motifs thereof. In still another embodiment the one or more features is the range of amounts of one or more carbohydrates (e.g., glycans). In yet another embodiment the one or more features is the range of ratios between two or more carbohydrates (e.g., glycans).

In one embodiment the glycoprofile is generated from determining the glycosylation site occupancy of glycoconjugates in a sample.

In another embodiment the glycoprofile is determined by identifying and quantifying all of the carbohydrates (e.g., glycans) and/or a monomer or motif thereof in the sample. In one embodiment the carbohydrates (e.g., glycans) are identified and quantified by solving constraints with a computational method.

5 In another aspect of the invention a method of generating a glycome pattern with a method as provided herein is also provided.

In still another aspect of the invention a method of creating a database of glycome patterns generated by a method provided herein is provided. In another embodiment the method also includes recording one or more values representing the glycome pattern in a
10 computer-generated data structure. The database so created is also provided.

Each of the limitations of the invention can encompass various embodiments of the invention. It is, therefore, anticipated that each of the limitations of the invention involving any one element or combinations of elements can be included in each aspect of the invention.

15

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 shows the conserved N-glycan pentasaccharide core.

Fig. 2 illustrates classes of N-linked glycans. High-mannose structures contain up to nine mannose residues (Fig. 2A). Complex type glycans are modified with hexosamines, galactoses, sialic acids and/or fucose, among other residues (Fig. 2B). Complex type chains
20 can occur as mono-, bi-, tri-, and tetra-antennary structures. Also, the amount and type of sialylation differs. Hybrid structures contain characteristics of both high-mannose and complex types (Fig. 2C).

Fig. 3 provides the detailed pathway of N-glycan biosynthesis (<http://www.genome.ad.jp/kegg/pathway/map/map00510.html>).

25 Fig. 4 shows the cleavage sites of Endo H, Endo F and PNGase F. Endo H can only act on high mannose and hybrid structures, while Endo F is effective at cleaving all classes of N-glycans. PNGase F also cleaves all mammalian N-glycan structures.

Fig. 5 provides the MALDI-MS spectra of N-glycans from RNaseB samples prepared by various methods. Glycans prepared using a GlycoClean S column (Table 2, Sample 12),
30 showed the expected high mannose peaks and significant contamination of unknown identity (Fig. 5A). A small amount of sample (10 μ g) was prepared using a 25 mg GlycoClean H column (Table 2, Sample 17), which showed only detergent peaks (Fig. 5B). A larger amount of protein (50 μ g) was prepared (Table 2, Sample 18), yielding the expected glycan

peaks but still containing detergent contamination (Fig. 5C). Using a 200 µg GlycoClean H column to purify N-glycans from 150 µg of RNaseB (Table 2, Sample 20), only the high mannose saccharides were observed (Fig. 5D).

Fig. 6 shows the spectrum from MALDI-MS of N-glycans from ovalbumin. Each labeled peak corresponds to a previously reported structure listed in Table 3.

Fig. 7 provides results from a study of N-glycans from antibody samples. Figs. 7A and 7B are for samples from Applikon bioreactors, with DO=50%, pH=7 and DO=90%, pH uncontrolled, respectively. Figs. 7C-7E are for samples from Wave reactors. Fig. 7C represents the results for DO controlled, pH uncontrolled, and NaOH in the media, while Fig. 7D represents the results with NaHCO₃ in the media instead of NaOH. The results shown in Fig. 7E are for DO uncontrolled with pH=7.

Fig. 8 shows the structures and theoretical masses of N-glycans released from antibodies.

Fig. 9 provides the MALDI-MS spectra of glycans released from serum proteins using PNGase F and Endo F. Serum samples were treated with PNGase F (Fig. 9A) or Endo F (Fig. 9B) and purified. While glycans were observed, the samples did not produce clean results. The peak cluster indicated by an arrow represents detergent contamination.

Fig. 10 shows a separation of neutral and acidic glycans using a GlycoClean H cartridge. Fig. 10A provides the results of an original mix of standards in positive mode. A3 and SC1840 are highly charged and do not ionize well. Fig. 10B shows that neutral glycans eluted off the GlycoClean H cartridge ionize well in positive mode, while only the charged sugars are present in the results provided by Fig. 10C, allowing them to be observed in negative mode. The multiple peaks in Fig. 10C arise from sodium adducts, typically one adduct per sialic acid residue.

Fig. 11 provides the results from MALDI-MS of N-glycans from human serum in neutral (left) and acidic (right) fractions. Figs. 11A and 11B show the results with neutral glycans prepared from two different IMPATH normal male human serum samples, while Figs. 11D and 11E show the results with the acidic fraction. Figs. 11C and 11F show the results with the neutral and acidic fractions of a normal human sample from Biomedical Resources.

Fig. 12 provides the results of serum glycans separated by ConA. Fig. 12A provides the results from SDS-PAGE of ConA flow-through (Lane 2) and elution (Lane 3). Lane 1 shows molecular weight standards. The results from MALDI-MS of neutral and acidic sugars obtained from ConA elution are shown in Figs. 12B and 12C.

Fig. 13 provides the results from Protein A separation of IgG from serum. A glycoblot of Protein A flow-through (Lane 3) and elution (Lane 4) is shown in Fig. 13A. Lane 1 contains protein standards, while Lane 2 (negative control) contains human serum albumin (* marks where albumin would run on an SDS-PAGE gel). Only glycosylated proteins are observed in the glycoblot, so the albumin does not stain. Figs. 13B (neutral) and 13C (acidic) show the results of MALDI-MS of glycans harvested from the elution fraction. Total serum glycans are pictured in Figs. 13D (neutral) and 13E (acidic).

Fig. 14 shows the permethylation of N-glycans. All OH and NH groups can be permethylated. For a complete reaction, it is important that the reaction vessel is free of air and water.

Fig. 15 shows the results of MALDI-MS of permethylated glycan standards. Fig. 15A shows that unmodified standards ionized unevenly. Fig. 15B shows that permethylated standards were more uniformly ionized, but generally did not have higher signal-to-noise ratios.

Fig. 16 shows the aminooxyacetyl peptide and its conjugation to N-glycans. The aminooxyacetate end of the synthetic peptide (top) reacts with the open form of the reducing end GlcNAc of N-glycans (bottom).

Fig. 17 shows the results of MALDI-MS of peptide-conjugated N-linked standards. Fig. 17A shows that unmodified glycans ionize unevenly, especially charged glycans (Figs. 17F and 17G). After conjugation with aminooxyacetyl peptide, ionization is much more uniform (Fig. 17B).

Fig. 18 shows the identification of serum N-glycans from MALDI-MS spectra. Fig. 18A shows the results of neutral glycans, while Fig. 18B shows the results of acidic glycans. Labeled peak numbers correspond to entries in Table 7.

Fig. 19 shows the results of neutral N-glycans from PVDF digest. Only the most abundant glycans are observed.

Fig. 20 provides MALDI spectra of glycans before (Fig. 20A) and after (Fig. 20B) applying a new recipe with optimized conditions.

Fig. 21 provides results from glycan quantification using an optimized matrix recipe for MALDI-MS.

Fig. 22 provides a schematic of an example of a methodology for analysis.

Fig. 23 provides a flowchart illustration of one example of a combined analytical-computational method for glycan analysis.

Fig. 24 provides a scheme for an exemplary method for glycoprotein analysis - glycan site occupancy analysis.

Fig. 25 provides results from a glycan site occupancy analysis for ribonuclease B. MS data for a peptide eluting at 7.8 minutes for unlabeled sample (**Fig. 25A**) and for the ¹⁶O/¹⁸O labeled 1:1 mixture (**Fig. 25B**) are provided. The expected [M+H]⁺ for the unlabeled peptide fragment is 476.29 Da.

Fig. 26 provides a MALDI-MS spectra of N-glycans from RNaseB with the expected high mannose structures.

Fig. 27 provides results from MALDI-MS of N-glycans from ovalbumin. Each labeled peak corresponds to a previously reported structure.

Fig. 28 provides structures and theoretical masses of N-glycans released from antibodies.

Fig. 29 shows the results from an analysis of depletion of serum albumin and IgGs from serum. **Fig. 29A** provides the results from a SDS gel stained with Simply Blue before (Lanes 7 and 14) and after removal of serum albumin and IgG using different conditions (Lanes 1-6 and 8-13). **Fig. 29B** provides the results from a Western blot (using Protein A-HRP detection) used for quantifying the removal of IgGs. Lanes 7 and 14 are without depletion, and Lanes 1-6 and 8-13 are using different conditions for the removal. **Fig. 29C** provides the quantification of IgG removal.

Fig. 30 shows the results of Protein A separation of IgG from serum. **Fig. 30A** provides the results from a glycoblot of Protein A flow-through (Lane 3) and elution (Lane 4). Lane 1 contains protein standards, while Lane 2 (negative control) contains human serum albumin (* marks where album would run on an SDS-PAGE gel). Only glycosylated proteins are observed in the glycoblot, so the albumin does not stain.

Fig. 31 shows the identification of serum N-glycans from MALDI-MS spectra. **Fig. 31A** shows the results of the neutral glycans, while **Fig. 31B** shows the results of the acidic glycans.

Fig. 32 provides the results from LC-MS (**Fig. 32A**) and CE-LIF (**Fig. 32B**) analysis of neutral glycome from serum.

Fig. 33 provides a MALDI-MS acidic glycome profile of saliva (**Fig. 33A**) and urine (**Fig. 33B**).

Fig. 34 provides a quantitative neutral glycome profile for serum with normal (**Fig. 34A**) and low (**Fig. 34B**) IgG levels.

Fig. 35 provides alterations in serum glycomic patterns between matched healthy (Fig. 35A) and cancer (Fig. 35B) patients.

Fig. 36 provides a schematic representation of an example of the computational strategy for the analysis of glycoprofile patterns.

Fig. 37 provides the results from a matrix comparison of MALDI-MS analysis of a hyaluronic acid fragment. DHB matrix (Fig. 37A); ATT-NafionTM matrix (Fig. 37B). Expected [M-H]⁻ = 4170.6.

Fig. 38 provides a schematic illustrating an exemplary analytic method using NMR, MALDI-MS and a computational approach.

Fig. 39 provides the structures satisfying experimental constraints. Mol. Wt. 1990; Man3Gal2Fuc1GlcNAc5 – 25% (Fig. 39A); Mol. Wt. 1990; Man3Gal2GlcNAc4GalNAc1Fuc1 – 50% (Fig. 39B); Mol. Wt. 2047; Man3Gal2GalNAc1GlcNAc5 – 25% (Fig. 39C).

Fig. 40 provides a schematic illustrating an exemplary method of glycoconjugate characterization.

DETAILED DESCRIPTION

It has been recognized that carbohydrates play a significant role in a variety of biological and pathological processes. However, information regarding which carbohydrates are important and how they affect biological functions is limited. Additional methods for analyzing carbohydrates are desirable. Such methods are provided herein and can be used for a number of purposes as described below.

Improved methods of analyzing carbohydrates are provided herein. Carbohydrates include, for example, starches, celluloses, gums and saccharides. Although, for illustration, the term “saccharide” or “glycan” is used below, this use is not intended to be limiting. It is intended that the methods provided herein can be directed to any carbohydrate, and the use of a specific kind of carbohydrate is merely exemplary.

As used herein, the term “saccharide” refers to a molecule comprising one or more monosaccharide groups. Saccharides, therefore, include mono-, di-, tri- and polysaccharides. A “polysaccharide”, as used herein, is any polymer made up of two or more monosaccharides consecutively linked through glycosidic linkages. Polysaccharides include those that are isolated from plant, animal and microbial (e.g., bacterial, viral) sources. The term “polysaccharide” as used herein, therefore, includes mucins, alginates, pectins, chitin,

pentosan, dextran sulfate, amylose, cellulose, etc. Polysaccharides also include glycosaminoglycans (GAGs), a family of complex polysaccharides that include dermatan sulfate (DS), chondroitin sulfate (CS), heparin, heparan sulfate (HS), keratan sulfate and hyaluronic acid (HA).

5 Polysaccharides further include glycans. Glycans, as used herein, are polysaccharides found on cells, proteins, lipids and in body fluids that are, generally, composed of hexoses, N-acetylhexosamines (HexNAcs), fucoses, sialic acids, etc. Each of these in turn can correspond to a single or multiple explicit monosaccharides, such as glucose (Glc), galactose (Gal), mannose (Man), N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc),
10 fucose (Fuc), N-acetylneuraminic acid (NeuAc), N-glycolylneuraminic acid (NeuGc), etc. Glycans can be branched or unbranched. The term "glycan" includes glycans that are intact (i.e., as they were originally found in nature or in a sample) or glycans that have been digested (i.e., fragment(s) of a glycan produced from chemical or enzymatic treatment.) The term is also intended to include charged and uncharged glycans and, therefore, neutral, acidic
15 and basic glycans.

Glycans can be found linked to non-saccharide components, such as lipids or proteins. Glycans linked to non-saccharide components are herein referred to as glycoconjugates. The term "glycoconjugate" refers to a conjugate of one or more glycans attached to a non-saccharide component. Generally, the attachment of the one or more glycans occurs through
20 covalent linkage. Glycoconjugates include glycoproteins, glycopeptides, peptidoglycans, proteoglycans, glycolipids and lipopolysaccharides. The exemplary use of any one of these terms is also not intended to be limiting. As used herein, a "peptide-based glycoconjugate" is meant to refer to glycoproteins, glycopeptides, peptidoglycans and proteoglycans, while a "lipid-based glycoconjugate" is meant to refer to glycolipids and lipopolysaccharides. As
25 used herein, "peptides" are intended to refer to proteins and polypeptides. Peptides, therefore, include short and long polypeptides as well as complete proteins.

Peptide-based glycoconjugates contain N- and O- glycans (also referred to herein as N- and O- linked glycans.) For illustration, but not intended to be limiting, N-glycans are generally classified into three types based on their structure: high mannose, hybrid and
30 complex (Sears, P., Wong, C.H. 1998. *Cell Mol Life Sci.* 54, 223-52.) All N-glycans contain a conserved pentasaccharide core composed of two N-acetylglucosamine residues followed by three mannose saccharides (Fig. 1). High mannose structures contain up to six more mannoses on both branches without further hexosamine, galactose or sialic acid residues (Fig. 2A), while complex structures have no additional mannoses on either arm (Fig. 2B).

Instead, they are composed of additional hexosamines and/or galactoses and/or sialic acids.

Hybrid structures are mixes of both high mannose and complex structures (Fig. 2C).

Additionally, branch termini can be capped with sialic acid (a charged monosaccharide), and the core or branches can be fucosylated. In addition, rare modifications exist, including sulfates, phosphates and xyloses, although these are typically not found in humans. The term “glycan” is intended to encompass these and other modified forms. O-glycans, on the other hand are assembled by series of reactions catalyzed by glycosyltransferases and sulfotransferases in the Golgi. In the O-glycan pathways, every sugar is transferred from a specific nucleotide sugar donor by the action of specific membrane-bound glycosyltransferases. In cancer cells, many of the enzymes involved in O-glycan biosynthesis are up- or down- regulated.

In addition to being found as part of a glycoconjugate, glycans can be found attached to the surface of a cell or they can be found in free form (i.e., separate from and not associated with a cell or other component.) In some instances, the glycans attached to the surface of a cell are one or more glycans that are part of a glycoconjugate, wherein the glycoconjugate is attached to or forms part of the cell's surface. Therefore, the methods provided herein can be used to analyze glycans that are part of glycoconjugates, attached to the surface of a cell, found in free form or some combination thereof. A “sample containing glycans” is meant to embrace a sample containing one or more glycans in any of these aforementioned forms. A “sample containing carbohydrates” is likewise meant to embrace a sample containing one or more carbohydrates in free form or as part of a complex or conjugate. As used herein, the “glycome” of a sample is all of the carbohydrates of the sample. The carbohydrates can be part of glycoconjugates but are not necessarily so.

It has been found that the analysis of carbohydrates (e.g., glycans), with an analytical (i.e., experimental) method in combination with a computational method results in improved analysis. Therefore, methods for analyzing samples containing carbohydrates (e.g., glycans) are provided, which include a combined analytical-computational platform. Non-limiting examples of such methods are illustrated in detail in the **Examples**.

In the methods provided herein, any analytical (or experimental) method for analyzing samples containing carbohydrates (e.g., glycans) so as to characterize them can be performed. As used herein, to “characterize” means to obtain data that can be used to determine the identity, structure, composition or quantity of a carbohydrate (e.g., glycan) or a glycoconjugate. The term also means to determine a property of the carbohydrates (e.g., glycans) or glycoconjugate. A “property” as used herein is a characteristic (e.g., structural

characteristic of the carbohydrates (e.g., glycans) or glycoconjugate that provides information about the carbohydrate (e.g., glycan) or glycoconjugate. Examples of properties include charge, chirality, nature of substituents (or components), quantity of substituents, molecular weight, molecular length, compositional ratios of substituents or units, type of basic building blocks (e.g., saccharide, amino acid, lipid constituents), hydrophobicity, enzymatic sensitivity, hydrophilicity, secondary structure and conformation, ratio of one set of modifications to another set of modifications, etc. When the term is used in reference to a glycoconjugate, it can also include determining the glycosylation sites, the glycosylation site occupancy, the identity, structure, composition or quantity of the carbohydrate (e.g., glycan) and/or non-saccharide component of the glycoconjugate as well as the identity and quantity of a specific glycoform.

As used herein, "glycosylation" is meant to include the pattern or a subset or even one particular carbohydrate (e.g., glycan), while "glycosylation pattern" refers to a pattern (or signature) that characterizes or distinguishes a sample with respect to the carbohydrates (e.g., glycans) present in the sample. A glycosylation pattern can be determined for a sample even if all of the details of the carbohydrate (e.g., glycan) structures are not known. A glycosylation pattern can provide, but is not required to, for example, the absolute or relative number, identity, etc. of the carbohydrates or components thereof in a sample. As used herein, a "component of a carbohydrate" is a monomer, set of monomers or a motif of the carbohydrate but is not the complete carbohydrate. As used herein, a motif of a carbohydrate is a specific set of monomers or subsequence of a carbohydrate. Generally, the motif includes 3, 4, 5 or more monomers of the carbohydrate. A "component of a glycoconjugate" includes the carbohydrate or portion thereof and the non-saccharide moiety or portion thereof.

In a population of glycoconjugates, each glycosylation site may (or may not) be occupied by a specific carbohydrate (e.g., glycan) all the time. Therefore, as used herein, a "glycoform" is a specific form of a glycoconjugate, which contains a particular carbohydrate (e.g., glycan) or set of carbohydrates (e.g., glycans) conjugated to a particular non-saccharide component at particular glycosylation site(s). As used herein, the term "glycosylation site occupancy" refers to the frequency (percentage) in which one or more specific glycosylation sites on a lipid or peptide is occupied by a carbohydrate (e.g., glycan). In one embodiment the glycosylation site occupancy is the "total glycosylation site occupancy", which refers to the frequencies in which all of the specific glycosylation sites on a lipid or peptide are occupied by a carbohydrate (e.g., glycan).

Analyzing glycoconjugates include analyzing the carbohydrates (e.g., glycans), the non-saccharide moieties, the complete glycoconjugate or some combination thereof. In some instances, the analysis of a glycoconjugate allows for the identification of the one or more carbohydrates (e.g., glycans) and the non-saccharide component of the glycoconjugate. The identification can, therefore, include determining the sequence of the one or more carbohydrates (e.g., glycans) and/or the non-saccharide component of the glycoconjugate. As an example, the characterization with an analytical method can include determining the peptide (or lipid) sequence, composition or structure of a peptide-based glycoconjugate (or lipid-based glycoconjugate). The analysis of a glycoconjugate can also include the determination of the glycosylation sites and/or the glycosylation site occupancy of the glycoconjugate. The specific carbohydrates (e.g., glycans) that occupy each specific glycosylation site can also be characterized using one or more analytical (or experimental) techniques. Analyses of glycoconjugates can, therefore, also include the identification and quantification of all of the glycoforms in a sample containing glycoconjugates. The method can further include the determination of the glycosylation sites and glycosylation site occupancy of the glycoconjugates.

In one aspect of the invention, therefore, a method is provided for the determination of the glycosylation sites and glycosylation site occupancy. In one embodiment this method includes cleaving the non-saccharide component of the glycoconjugates, cleaving the carbohydrates (e.g., glycans) from the non-saccharide components, labeling the non-saccharide components of a first portion of the sample at the glycosylation sites, cleaving the carbohydrates (e.g., glycans) from the non-saccharide components of a second portion of the sample, analyzing the first and second portions of the sample containing the non-saccharide components and comparing the results. The glycosylation site occupancy can be quantified from ratios of the masses of the non-saccharide component of the first and second portions of the sample. The non-saccharide components of the first portion of the sample can be labeled with a labeling agent, and the non-saccharide components of the second portion of the sample can be labeled or unlabeled.

Labeling agents as used herein include isotopes of C, N, H, S or O. In one particular example the labeling agent is an isotope of O, such as ^{18}O . In another example the labeling agent is ^2H .

Carbohydrates (e.g., glycans), glycoconjugates and non-saccharide components can be analyzed using any of a number of analytical methods. Analytical methods include, for example, mass spectrometric methods, nuclear magnetic resonance (NMR), electrophoretic

methods and chromatographic methods. Examples of mass spectrometric methods include ESI-MS, LC-MS, LC-MS/MS, MALDI-MS, MALDI-MS/MS, MALDI-TOF-MS, MALDI/PSD-MS, MALDI-TOF/TOF-MS, MALDI-FTMS, LC-MALDI-MS, LC-MALDI-TOF-TOF-MS, Nano-LC MALDI-TOF-TOF-MS, TANDEM-MS, etc.

- 5 Analysis of a sample containing carbohydrates (e.g., glycans) (e.g., with a mass spectrometric method) can, for example, provide information regarding the monomer composition and/or their relative abundance. As used herein, "monomer composition" refers to the identity and/or quantity of the monomers that make up a carbohydrate. When the carbohydrate is a polysaccharide, such as a glycan, the term is "monosaccharide
- 10 composition" (e.g., the number of hexoses, N-acetyl hexosamines, fucoses, sialic acids, etc.) "Relative abundance of the monomers" refers to the ratios of the relative amounts of particular monomers to other monomers.

- In some embodiments, the analytical method includes the use of MALDI-MS. Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) techniques for
- 15 the analysis of oligosaccharides have been described (Juhasz, P. & Biemann, K. (1995) *Carbohydr Res* 270, 131-47 and Juhasz, P. & Biemann, K. (1994) *Proc Natl Acad Sci U S A* 91, 4333-7; Venkataraman, G., Shriver, Z., Raman, R. & Sasisekharan, R. (1999) *Science* 286, 537-42; Rhomberg, A. J., Shriver, Z., Biemann, K. & Sasisekharan, R. (1998) *Proc Natl Acad Sci U S A* 95, 12232-7; Ernst, S., Rhomberg, A. J., Biemann, K. & Sasisekharan, R.
- 20 (1998) *Proc Natl Acad Sci U S A* 95, 4182-7; and Rhomberg, A. J., Ernst, S., Sasisekharan, R. & Biemann, K. (1998) *Proc Natl Acad Sci U S A* 95, 4176-81). Optimized MALDI-MS analytical methods are also provided herein.

- NMR methods include, for example, simple 1D, 2D, COSY, gCOSY, TOCSY, NOESY, etc. When NMR is used to analyze a sample containing carbohydrates (e.g.,
- 25 glycans), the results from the NMR analysis can provide information regarding the monomer composition and/or linkage for the carbohydrates (e.g., glycans). "Linkage information" refers to the type and/or abundance of particular linkages between monomers (or monosaccharides in the case of polysaccharides). "Linkage abundance" is used to refer to the absolute or relative amounts (i.e., as ratios) of particular linkages. Types of linkages present
- 30 in glycans include, for example, NeuNAc α 2-3Gal, NeuNAc α 2-6Gal, GlcNAc β 1-2Man, GlcNAc β 1-4Man, Man α 1-6Man, Man α 1-3Man, Man α 1-2Man, Gal β 1-3GalNAc, GlcNAc β 1-6GalNAc, GlcNAc β 1-3GalNAc, etc. NMR analysis can also provide information regarding the ratios of the monomers of the carbohydrates (e.g., glycans). NMR can also be

used to determine the glycosylation site occupancy of a glycoconjugate. NMR can further be used to determine monomer composition as well as relative amount (i.e., ratios of particular sets of monomers).

As an example, 2D-NMR can be used for the identification of N-linked and O-linked glycan site occupancy. A combination of COSY, TOCSY, NOESY experiments can be first conducted on a specific quantity of a peptide-based glycoconjugate. Using COSY and TOCSY data, all the spin systems (amino acids) can be assigned. NOESY experiments can also be used to determine the specific amino acid sequence. This information allows the specific identification of all the asparagines (Asn) and serine (Ser) or threonine (Thr) residues in the sample. NOEs between the protons of the Asn, Ser or Thr side chains and proximal carbohydrate residues can be easily monitored, which allow the monitoring and quantification of carbohydrate (e.g., glycan) occupancy at each glycosylation site. This is particularly useful for high abundance glycosylation sites.

Incorporating NMR data as constraints to further refine mass spectrometric information (e.g., MALDI-MS information) enables the elimination of explicit compositions that do not satisfy the monomer (e.g., monosaccharide) composition data and a more quantitative determination of the abundance of monomers (e.g., monosaccharide) and linkage distributions. In addition, biosynthetic rules and database look-ups (e.g., <http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp>) can help in further convergence of the solution to obtain an accurate picture of the number and relative abundance of the species in the sample as well as the best characterization of the individual structures corresponding to these species. Important NMR information that can be used as constraints to refine the carbohydrate (e.g., glycan) structures are, for example, the linkage abundance between certain monomers (e.g., monosaccharides) and/or the specific ratios between them. Examples of these include Man α 1-6Man, Man β 1-4GlcNAc, GlcNAc β 1-4GlcNAc, Man α 1-3Man, GlcNAc β 1-6Man, GlcNAc β 1-4Man, GlcNAc β 1-2Man, Gal β 1-4GlcNAc, Gal α 1-3Gal, Fuc α 1-6GlcNAc, GalNAc β 1-4GlcNAc, NeuNAc α 2-3Gal, NeuNAc α 2-6Gal, Man α 1-2Man, Gal β 1-3GalNAc, GlcNAc β 1-6GalNAc and GlcNAc β 1-3GalNAc. Methods using such information are also provided herein.

Electrophoretic methods include, for example, gel electrophoresis, capillary electrophoresis (CE) and capillary electrophoresis-laser induced fluorescence (CE-LIF), etc. Some of the electrophoretic methods can further include the labeling of the carbohydrates

(e.g., glycans) with fluorophores, such as, for example, fluorescence-assisted carbohydrate electrophoresis (FACE) and CE-LIF. A method for the compositional analysis of oligosaccharides using CE has been described (Rhomberg, A. J., Ernst, S., Sasisekharan, R. & Biemann, K. (1998) *Proc Natl Acad Sci U S A* 95, 4176-81).

5 Chromatographic methods include high performance liquid chromatography (HPLC).

 Samples containing carbohydrates (e.g., glycans) can be analyzed with any of the analytical methods provided herein. The methods can further include a step of contacting a sample containing carbohydrates (e.g., glycans) with acidic or basic conditions in order to cleave the carbohydrates (e.g., glycans) or monomers (e.g., monosaccharides) from the
10 carbohydrates (e.g., glycans). For example, glycans can be cleaved by treating a sample with hydrazine or basic borohydride. Sialic acid residues can be cleaved using acidic conditions and high temperature (for example, sulfuric acid at 80° C).

 Carbohydrates (e.g., glycans) can also be quantified using calibration curves of known carbohydrate (e.g., glycan) standards. More detailed examples of these methods are provide
15 below in the **Examples**.

 Any of the analytical methods provided herein can further comprise the use of carbohydrate- or glycan-degrading enzymes, such as by contacting a sample containing glycans with a glycan-degrading enzyme. As used herein "carbohydrate-degrading enzymes" or "glycan-degrading enzymes" are enzymes that modify a carbohydrate or glycan,
20 respectively, in some way. As one example, the modification can be the cleavage of the carbohydrate or glycan. Following enzymatic degradation, a sample of degraded carbohydrates (e.g., glycans) can be analyzed with a method as described herein. Examples of glycan-degrading enzymes are known in the art and include sialidase, galactosidase, mannosidase, N-acetylhexosaminidase or a combination thereof.

25 The information gathered from the analytical methods can be used to generate constraints. As an example, a method of analyzing a sample containing carbohydrates (e.g., glycans) with analytical and computational methods can include the steps of analyzing the sample with an analytical method (e.g., performing an experiment on the sample), obtaining results, generating constraints from the results, and solving the constraints. As used herein, a
30 "computational method" is any method that involves establishing and/or solving a logic expression (e.g., mathematical relationship). "Constraints", as used herein, are relationships of one or more values, results or information about a sample containing carbohydrates (e.g., glycans) that can be compared or evaluated as part of a computational method. The relationships can be mathematical equations and/or equalities (e.g., equal to, at most, at least,

including, etc.) The constraints can, for example, be one or more mathematical equations generated with the data obtained from an analysis of a sample containing carbohydrates (e.g., glycans) and/or other information obtained from other sources, such as databases or with other analytical methods. As part of a computational method, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 5 15 or more logic expressions (e.g., mathematical equations) can be generated.

As described above, constraints can be generated with data obtained from the results of an analytical method. Therefore, constraints can be generated from the results of an analysis of a glycoconjugate, a non-saccharide component of a glycoconjugate, a carbohydrate (e.g., glycan) or some combination thereof. Constraints can also be generated from information regarding a component of a carbohydrate (e.g., glycan) and/or a glycoconjugate. Constraints can also be generated with other information, such as information from databases that contain information about carbohydrates (e.g., glycans), glycoconjugates and/or non-saccharide components, as well as information regarding mass, enzyme action and/or biosynthesis. The databases referred to herein can be those described herein in the **Examples**, known in the art or can be generated with the methods provided. Constraints can also be generated from information regarding the origin of a sample, the expression system or expression conditions for the synthesis of a carbohydrate (e.g., glycan) or glycoconjugate, the species from which a carbohydrate (e.g., glycan) or glycoconjugate is derived, the expression levels of glycosidases and glycosyltransferases from the source of a carbohydrate (e.g., glycan) or glycoconjugate or the state of the source of a carbohydrate (e.g., glycan) or glycoconjugate. 10 15 20

As mentioned above, the constraints can be generated using, for instance, what is known of the biosynthetic pathway of glycan synthesis. Unlike DNA or protein synthesis, which are template-driven processes, glycan biosynthesis is a complex process involving a multitude of enzymes. A detailed scheme of N-glycan biosynthesis is shown in **Fig. 3** and the biosynthetic enzymes and their EC numbers are listed in **Table 1**. The process is initiated in the cytoplasm, with the nascent sugar attached to the endoplasmic reticulum (ER) membrane through a lipid anchor. After a glycan core of two glucosamines followed by five mannose residues is constructed, the orientation of the growing glycan is flipped to face the lumen of the ER. There, four more mannose residues are added by α -mannosyltransferase, and one branch is capped with three glucoses. At this point, oligosaccharyl transferase catalyzes the removal of the naïve glycan from its lipid anchor and attaches it to a glycosylation site on a protein undergoing synthesis in the ER (Varki, A. (1999) *Essentials of glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.) 25 30

To ensure that the glycan can play its proper role in protein folding and transport, the three terminal glucose residues and one mannose are removed. This trimming is required for the glycan to interact with the chaperone proteins calnexin and calreticulin (Helenius, A., Aebi, M. (2001) *Science* 291, 2364-9; Parodi, A.J. (2000) *Annu Rev Biochem* 69, 69-93.) As the correctly folded protein passes through the Golgi on its way either to secretion or the cell membrane, further glycan modifications can take place. Specifically, mannosidases can trim more mannoses off the core sugar, while a host of glycosyltransferases can add further GlcNAc, fucose, galactose and sialic acid moieties, among others (Sears, P., Wong, C.H. (1998) *Cell Mol Life Sci* 54, 223-52.)

10

Table 1. Common Enzymes Involved in N-glycan Biosynthesis

EC #	Enzyme name	EC #	Enzyme name
2.4.1.-	Hexosyltransferases (ALG 6, 8, 10, 11)	2.4.1.155	α -1,6-mannosyl-glycoprotein 6- β -N-acetylglucosaminyl transferase
2.4.1.38	Glycoprotein β -galactosyltransferase	2.4.1.201	β -1,6-mannosyl-glycoprotein 4- β -N-acetylglucosaminyl transferase
2.4.1.68	Glycoprotein 6- α -L-fucosyltransferase	2.4.99.1	β -galactoside α -2,6-sialyltransferase
2.4.1.83	Dolichyl phosphate mannose transferase	2.5.1.-	Transferring alkyl or aryl groups, other than methyl groups
2.4.1.101	α -1,3-mannosyl-glycoprotein 2- β -N-acetylglucosaminyl transferase	2.7.1.108	Dolichol kinase
2.4.1.117	Dolichyl phosphate β -glucosyltransferase	2.7.8.15	Chitobiosylpyrophosphoryl dolichol synthase
2.4.1.119	Oligomannosyl transferase	3.1.3.51	Dolichyl-phosphatase
2.4.1.130	Oligomannosyl synthase (ALG 3, 9, 12)	3.1.4.48	dolichylphosphate-glucose phosphodiesterase
2.4.1.132	Glycolipid 3- α -mannosyltransferase	3.2.1.-	Hydrolyzing O- and S-glycosyl compounds
2.4.1.141	N,N'-diacetylchitobiosyl pyrophosphoryldolichol synthase	3.2.1.106	Mannosyl-oligosaccharide glucosidase
2.4.1.142	chitobiosyldiphosphodolichol β -mannosyltransferase	3.2.1.113	mannosyl-oligosaccharide 1,2- α -mannosidase
2.4.1.143	α -1,6-mannosyl-glycoprotein 2- β -N-acetylglucosaminyl transferase	3.2.1.114	mannosyl-oligosaccharide 1,3-1,6- α -mannosidase
2.4.1.144	β -1,4-mannosyl-glycoprotein 4- β -N-acetylglucosaminyl transferase	3.6.1.43	Dolichol diphosphatase
2.4.1.145	α -1,3-mannosyl-glycoprotein 4- β -N-acetylglucosaminyl transferase		

Constraints can be solved using mathematical and heuristic approaches known in the art based on the specific constraints generated for a specific problem. The approaches can range from standard numerical methods, such as Gaussian Elimination to more complex methods, such as linear programming and simulated annealing. Other approaches that may be used to solve the constraints include parameter estimation approaches, such as least squares and non-linear methods. Yet another class of approaches are those based on search techniques that generate optimal solutions. Many of these mathematical and heuristic methods are available as computer programs and mathematical software.

A detailed example of an analytical and computational method is provided in Fig. 23. One of skill in the art will appreciate, however, that there are a number of ways in which analytical methods can be combined with computational methods to achieve the desired characterization of a sample containing carbohydrates (e.g., glycans). In some embodiments it is the combination which provides more efficient analysis. The examples provided are not intended to be limiting.

It has also been found that the combination of mass spectrometry (MS), such as MALDI-MS, and NMR also provides for the improved analysis of carbohydrates. When a sample containing carbohydrates (e.g., glycans) is analyzed, methods using MS and NMR, in one embodiment, can allow for the simultaneous assignment of monomer composition, linkages between monomers and detailed information about the chemical structure of the carbohydrates (e.g., glycans) in the sample.

In one example, MALDI-MS can provide the molecular weight of each glycan in a sample in one single profile of mass to charge ratio vs. intensity of the peak. Typically this gives the molecular mass since the charge observed in MALDI-MS is 1. Furthermore, depending on the mode of operation of the MALDI-MS instrumentation, negatively charged glycans can be analyzed distinctly from neutral glycans. Based on the molecular weight, the specific composition of the glycan in terms of number of hexoses, N-acetyl hexosamines, fucoses and sialic acids can be obtained. The data can, therefore, provide a first set of constraints for a computational method. In addition to providing the distinct mass signature for each glycan in the mixture, the MALDI-MS technique can also be optimized to provide quantitative information on the relative abundance of the glycans. This information can also be used as a constraint that provides a boundary for a computational method to assign the exact glycan structures for each mass peak.

NMR can be used in combination with the above-described MALDI-MS analysis to provide additional information to further characterize the glycans of the sample. Since, for example, the hexoses can include galactoses, mannoses and glucoses, and the HexNAcs can include N-acetylglucosamines and N-acetylgalactosamines. The anomeric proton and carbon of each monosaccharide in a glycan has a distinct chemical shift and thus provides a signature for quantifying each monosaccharide in a glycan mixture. Thus, the 1D proton of a glycan mixture along with the coupling constants, which can further be determined using gCOSY and TOCSY, can provide quantitative information about distinct monosaccharides in the mixture. For example, the ratios in the abundance (ratios of the absolute or relative amounts) of glucose to galactose to mannose can be obtained. These parameters can be lumped into hexose abundance in the MALDI-MS data. Thus, using the explicit monosaccharide composition based on NMR can provide another constraint for a computational framework to assign the glycan structures in the mixture.

In addition to the monosaccharide composition, NMR spectroscopy can also provide, for example, quantitative information on the linkages between monosaccharides. This information is important, for example, for terminal sialic acids which can be α 2-3 or α 2-6 linked to the penultimate monosaccharide. This linkage cannot be explicitly assigned using MALDI-MS data. The anomeric chemical shifts of the monosaccharides can further be classified based on the neighboring monosaccharide (at the reducing end), which can provide the abundance of the specific linkage between the two monosaccharides. The linkage abundance is important, since it is required to completely assign the glycan structure. While sample amounts have been a limiting factor for complete structure assignment of glycans using NMR spectroscopy, simple 1D proton and 2D gCOSY experiments do not require as much sample but can provide much information about a sample containing glycans. These experiments can also be more sensitive with low sample amounts compared to NOESY experiments, therefore, in some embodiments, 1D and 2D gCOSY analytical methods may be preferred.

In addition to the MALDI-MS and NMR analysis, a computational method can be used to incorporate the MALDI-MS and NMR data as constraints. There are multiple ways to develop the computational method for incorporating the analytical data as constraints and in searching for a solution of the constraints (i.e., obtaining the most accurate chemical structure information for the carbohydrates (e.g., glycans) in the sample.) In the case of N-linked glycans, for example, although the biosynthesis is complex, it is well known in terms of an ordered set of events which lead to the diversity of glycans. Thus, this knowledge of

biosynthesis can be encoded as rules to construct the entire solution space of theoretically possible glycan structures based on the mass and composition information obtained from the MALDI-MS data set. This large solution space can be narrowed during each step of applying other information such as relative abundance between two glycans, explicit monosaccharide composition, linkage abundance, etc. as constraints to give the final best solution in terms of the chemical structures of glycans in the sample. In the case of O-linked glycans, the biosynthesis rules are less defined, thus starting from a theoretical solution space of all possibilities might be cumbersome. For O-linked glycans, therefore, although not required, a heuristic approach of constructing the most appropriate solution space based on monosaccharide composition and linkage abundance can be used to provide a rapid way of identifying the solution.

In some embodiments, the methods provided herein also include generating a list of the possible compositions of carbohydrates (e.g., glycans), glycoconjugates or components thereof and their theoretical masses. The list can be based on the biosynthetic pathways for glycosylation (Fig. 3). The list can be generated with other means, such as with the results from the use of any of the methods provided herein or known in the art. The methods provided can also comprise or consist of the steps of generating a list of carbohydrate (e.g., glycan) or glycoconjugate properties. One example of such a method includes measuring 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more properties of a carbohydrate (e.g., glycan) or glycoconjugate and recording a value for the one or more properties to generate a list of carbohydrate (e.g., glycan) or glycoconjugate properties. In one embodiment the list comprises the number of one or more types of monomers (e.g., monosaccharides). The list can also include the total mass of a carbohydrate (e.g., glycan) or glycoconjugate, the mass of a non-saccharide component of a glycoconjugate, the mass of a carbohydrate (e.g., glycan), etc.

The list in one embodiment can be a data structure tangibly embodied in a computer-readable medium, such as computer hard drive, floppy disk, CD-ROM, etc. Table 6 represents an example of such a list. The list of Table 6 has a plurality of entries, where each entry encodes a value of a property. The values encoded can be any kind of value, such as, for example, single-bit values, single-digit hexadecimal values or decimal values, etc.

Therefore, also provided herein is a database, tangibly embodied in a computer-readable medium, wherein the database stores information descriptive of one or more carbohydrates (e.g., glycans) and/or glycoconjugates. The database comprises data units that correspond to the carbohydrate (e.g., glycan) and/or glycoconjugate. The data units include an identifier that includes one or more fields, each field storing a value corresponding to one

or more properties of the carbohydrates (e.g., glycans) and/or glycoconjugates. In one embodiment the identifier includes 2, 3, 4, 5, 6, 7, 8, 9, 10 or more fields. The database, for example, can be a database of all possible glycoconjugates and/or carbohydrates (e.g., glycans) or can be a database of values representing a glycome profile or pattern for one or more samples. Methods of analyzing and/or determining a glycome profile or pattern are described further below.

Carbohydrates (e.g., glycans) can be charged or uncharged. They can be acidic, basic or neutral. It has also been found that separately analyzing charged and uncharged carbohydrates (e.g., glycans) of a sample can provide an improvement in the analysis of carbohydrates (e.g., glycans). Therefore, the charged and uncharged carbohydrates (e.g., glycans) can be separated prior to the analysis of the carbohydrates (e.g., glycans), such as with an analytical method or other method as provided herein. As described further in the **Examples** provided below, such a method has been found to discriminate the carbohydrates (e.g., glycans) present in the sample. Therefore, any of the methods provided herein can include a step of separating neutral and charged carbohydrates (e.g., glycans), such as acidic carbohydrates (e.g., glycans).

Such separation can be achieved using purification methods. For instance, in a preferred embodiment, the separation is accomplished with a graphitic carbon purification cartridge by eluting glycan pools with different concentrations of acetonitrile. Other methods will be known to those of skill in the art. Analysis of these separate glycan pools can then be undertaken. For instance, when using MALDI-MS, acidic glycans can be analyzed in negative ion mode, while neutral glycans can be analyzed in positive ion mode.

In some analytical methods, such as in the analysis with MALDI-MS, the matrix in which the sample containing carbohydrates (e.g., glycans) is suspended can affect the quality of the analysis. It has been found that analysis of a sample containing carbohydrates (e.g., glycans) is improved when the sample is analyzed in the presence of a thymine derivative and an ion exchange resin. The thymine derivative can be thiothymine, 2-thiothymine, 4-thiothymine, 5-aza-2-thiothymine or 6-aza-2-thiothymine (ATT). The ion exchange resin can be an ammonium resin, a cationic exchange resin, a cationic exchange resin in pyridinium form, an anionic exchange resin or a perfluorinated ion exchange resin. The perfluorinated ion exchange resin can be, for example, NafionTM.

Other matrices can also result in improved analysis. In some embodiments the matrix preparation is caffeic acid with or without spermine. In other embodiments, the matrix preparation is dihydroxybenzoic acid (DHB) with or without spermine. In still other

embodiments the matrix preparation is spermine with DHB. The spermine, for example, can be in the matrix preparation at a concentration of 300 mM. The matrix preparation can also be a combination of DHB, spermine and acetonitrile. Additionally, the matrix can be a mixture of 5-methoxysalicylic acid (5-MSA) and DHB.

5 Additionally, instrument parameters can also be modified. These parameters may include guide wire voltage, accelerating voltage, grid values and negative versus positive polarity. In other embodiments, spot morphology can be employed to improve signal intensity.

 In general, when the carbohydrates (e.g., glycans) in a sample are part of a
10 glycoconjugate, the sample of glycoconjugates can be first denatured with a denaturing agent. A "denaturing agent" is an agent that alters the structure of a molecule, such as a protein. Denaturing agents, therefore, include agents that cause a molecule, such as a protein to unfold. Denaturing agents include those that comprise detergents, urea, high salt concentration, guanidium hydrochloride or heat. The denaturation can be followed by
15 reduction, which can be followed by carboxymethylation (or alkylation), etc. Reduction can be accomplished with a reducing agent, such as, dithiothreitol (DTT), β -mercapto ethanol (BME) or tris(2-carboxyethyl)phosphine (TCEP). Carboxymethylation or alkylation can be accomplished with, for example, iodoacetic acid or iodoacetamide. In some methods, therefore, following denaturation the sample of glycoconjugates is reduced with a reducing
20 agent. In other embodiments the sample of glycoconjugates is alkylated after being reduced.

 The methods provided herein can also include cleaving carbohydrates (e.g., glycans) from a non-saccharide component using any chemical or enzymatic method or combination thereof known in the art. In one embodiment this occurs prior to analysis. An example of a chemical method for cleaving is treatment of glycoconjugates with hydrazine or alkali
25 borohydride. Enzymatic methods include the use of enzymes specific to N- or O- linked sugars. These enzymatic methods, therefore, include the use of endoglycosidase H (Endo H), Endo F, N-Glycanase F (PNGase F) or a combination thereof. In some embodiments, PNGase F is used when the release of N-glycans is desired. When PNGase F is used for glycan release from a peptide-based glycoconjugate, the protein is, in some embodiments,
30 unfolded prior to the use of the enzyme. The unfolding of the protein can be accomplished with denaturation as provided above.

 After the release of the carbohydrate (e.g., glycan) from the non-saccharide component, or when the carbohydrates (e.g., glycans) of a sample are in free form (not part of a glycoconjugate), the sample containing carbohydrates (e.g., glycans) can be purified, for

instance, by precipitating the proteins with ethanol and removing the supernatant containing the carbohydrates (e.g., glycans). Other experimental methods for removing the proteins, detergent (from a denaturing step) and salts include methods known in the art. These methods include dialysis, chromatographic methods, etc. In one example, the purification is accomplished with a solid phase extraction cartridge or ion exchange resin. The solid phase extraction column can be, for example, a graphitic carbon column, non-graphitic carbon column or C-18 column.

Samples can also be purified with commercially available resins and cartridges for clean-up after chemical cleavage or enzymatic digestion used to separate carbohydrates (e.g., glycans) from the non-saccharide components. Such resins and cartridges include ion exchange resins and purification columns, such as GlycoClean H, S, and R (Glyco H, Glyco S and Glyco R, respectively) cartridges. In some embodiments GlycoClean H is used for purification. In still other embodiments, everything but the carbohydrates (e.g., glycans) are removed from the sample.

Purification can also include the removal of high abundance proteins, such as the removal of albumin and/or antibodies, from a sample containing carbohydrates (e.g., glycans). In some methods the purification can also include the removal of unglycosylated molecules, such as unglycosylated proteins. Removal of high abundance proteins can be a desirable step in some methods, such as some high-throughput methods (described elsewhere herein). In some embodiments, abundant proteins, such as albumin and/or antibodies, can be removed from the samples prior to the final analysis of a sample containing carbohydrates (e.g., glycans).

Prior to the analysis of a sample containing carbohydrates (e.g., glycans), the sample can be fractionated. The sample can be fractionated in order to obtain a sample of carbohydrates (e.g., glycans) that are a specific subgroup of molecules. "Subgroups of molecules" include molecules of specific properties, such as charge, molecular weight, size, binding properties to other molecules or materials, acidity, basicity, pI, hydrophobicity, hydrophilicity, etc. In one embodiment the subgroup of molecules is a low abundance glycan species (or group thereof), and it is the low abundance glycan species (or group thereof) that is analyzed with the methods provided. Low abundance glycan species include, but are not limited to, glycans that can contain fucoses, sialic acids, galactoses, mannoses or sulfate groups. In another embodiment the subgroup of molecules is a group of high abundance proteins. In one embodiment the subgroup of molecules is, therefore, the antibodies of a

sample. Therefore, the methods provided herein can be used for the analysis of the carbohydrates (e.g., glycans) of a subgroup of molecules.

A sample can be fractionated based on properties of the carbohydrates (e.g., glycans) and/or glycoconjugates, such as but not limited to, charge, size, molecular weight, binding properties to other molecules or materials, acidity, basicity, pI, hydrophobicity and hydrophilicity. As an example, the fractionation can be performed using solid supports with immobilized proteins, organic molecules, inorganic molecules, lipids, carbohydrates, nucleic acids, etc. As a further example, the fractionation can be performed using filters, such as molecular weight cutoff (MWCO) filters. The fractionation can also be performed using resins, such as, cationic or anionic exchange resins, etc. Any method of fractionation known in the art can be used. In one embodiment, however, the sample is not fractionated before it is analyzed by an analytical method as provided herein.

In other embodiments, carbohydrates (e.g., glycans) can be modified to improve their ionization, such as when MALDI-MS is used for analysis. Such modifications include permethylation and conjugation of a glycan to a peptide or derivitization with an organic molecule such as a chromophore. In other embodiments, the carbohydrates (e.g., glycans) are not modified prior to their analysis.

Samples of carbohydrates (e.g., glycans) can be analyzed separately, or they can be analyzed as a mixture. Therefore, samples containing carbohydrates (e.g., glycans) can be analyzed by first separation of a sample into portions of the sample and analyzing the portions separately or in some combination. The methods provided include methods for the analysis of the glycosylation of a single glycoconjugate or a mixture of glycoconjugates in a sample. Such mixtures can contain glycosylated and non-glycosylated peptides and/or lipids.

The methods provided herein can have a limit of detection of less than 1000, 500, 100, 75, 50, 25, 20, 18, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 femtomole.

The carbohydrates (e.g., glycans) or glycoconjugates that are analyzed with the methods provided herein can be analyzed in solution or when immobilized on a solid support. In one embodiment the solid support is in a 96-well plate format. In another embodiment the solid support is an individual membrane. In yet another embodiment the solid support can be in a 96-well plate format that comprises a membrane. Membranes, as used herein, include protein-binding membranes, such as polyvinylidene difluoride (PVDF) membranes, C-18 membranes and nitrocellulose membranes.

The methods provided herein can be performed in a high-throughput manner. "High-throughput" methods refer to the ability to process and/or analyze multiple samples at one

time. High-throughput methods provided herein can include the use of a membrane-based method, such as with a protein-binding membrane. High-throughput methods can also be performed in a 96-well plate format. In one embodiment the 96-well plate contains a protein-binding membrane. The methods provided, therefore, can include high-throughput sample processing steps (i.e., purification, digestion and/or denaturation steps, etc. performed in a high-throughput manner). Any step or steps of any of the methods provided herein can be performed in a high-throughput manner. In some embodiments protein-binding membrane based high-throughput methods can also include the removal of abundant proteins such as albumin.

The methods provided can also include the use of robotics. Robotics can be used in, for example, denaturation, reduction, alkylation, purification and fractionation steps.

Building on the description above, also provided in one aspect of the invention is a method for determining glycosylation site occupancy of a glycoconjugate. For the determination of the glycan site occupancy, such as for lower abundance glycoforms, concepts from phosphoproteomics were adapted. Fig. 24 provides one embodiment of a method for determining glycosylation site occupancy. Briefly, a well-characterized batch of the glycoprotein under study is used to generate a library of labeled peptides and glycopeptides by protease digest. In order to facilitate the determination of the glycosylation sites, each glycosylated amino acid can be differentially labeled. The labels that can be used include isotopes of C, N, H, S or O. In one embodiment the glycosylated amino acids are labeled with ^{18}O or unlabeled (^{16}O) using methods known in the art (Kaji, 2003). After the labeling, the samples can be further analyzed. In one embodiment the glycan site occupancy is quantified from the ratios of the masses of the labeled and unlabeled fragments. In one example, determining the glycosylation site and its occupancy can include cleaving and labeling with a first label the glycoconjugates at the glycosylation sites of a portion of the sample, cleaving the glycoconjugates at the glycosylation sites of another portion of the sample and analyzing the portion of the sample. In one embodiment the glycosylation sites of both portions of the sample are labeled. The portions of the sample can be analyzed separately or as a mixture in any ratio. First instance, when there are two portions of the sample, the two portions can be mixed in a 1:1, 1:2, 1:3, 1:4 or 1:5 ratio. The glycosylation site occupancy method can be used to determine the identity and number of glycoforms in the sample. Therefore, a method of determining the identity and number of glycoforms in a sample comprising determining the glycosylation site occupancy of a glycoconjugate and

analysis to characterize the glycoconjugates so as to determine the identity and number of glycoforms is also provided.

As illustrated in the **Examples** below, a fragment containing a partner peak with a molecular weight 2Da heavier was identified as a peptide containing a glycosylation site. By comparing the data between the glycosylated and deglycosylated samples, a preliminary identification all the peptides (or lipids when the glycoconjugate is lipid-based) and glycopeptides (or glycolipids) can be identified, and a preliminary identification of the carbohydrates (e.g., glycans) can be obtained. This quantitative information can be combined with an analytical method and used as constraints in a computational method to arrive at the complete characterization of the glycoconjugate.

Also provided is a method of generating a library. The library, in this instance, consists of labeled glycoconjugates and fragments of the glycoconjugates, the fragments being the non-saccharide components of the glycoconjugates. In one example, a library is generated by cleaving the backbone of the glycoconjugate and labeling the non-saccharide components of the glycoconjugates that result with a labeling agent. This example also includes the step of cleaving the carbohydrates (e.g., glycans) from a glycoconjugate. The carbohydrates (e.g., glycans) can then be removed from the sample. The libraries so produced can be analyzed with the methods provided herein. The libraries can also be used as a standard once characterized and methods of using such libraries are also provided. In one example, a method of analyzing a sample containing glycoconjugates includes cleaving the glycoconjugates, enzymatically removing the carbohydrates (e.g., glycans) from the glycoconjugates and mixing the sample with a standard. The sample mixed with the standard can then be analyzed. In one embodiment the amounts of the glycoconjugates and non-saccharide components of the sample and standard are compared. In one aspect of the invention methods of producing such standards are also provided.

Protein glycosylation can affect the function of proteins or be indicative of a cause or symptom of a disease state. For many proteins, N- and O- linked glycans are important factors for determining proper folding, stability and resistance to degradation (which affects the half-life of the protein). In some proteins, N- and O- linked glycans play a role in the activity and/or function of the protein. In some proteins, N- and O- linked glycans are indicative of a normal or disease state.

The methods provided, therefore, are also directed to the analysis of the total glycome of a sample. Such methods include the steps of analyzing the carbohydrates (e.g., glycans) of the sample and determining the profile of the carbohydrates (e.g., glycans). The

carbohydrates (e.g., glycans) can be analyzed with any of the methods provided herein. The "total glycome" refers to all of the carbohydrates (e.g., glycans) that can found in a sample. For instance, the carbohydrates (e.g., glycans) of the total glycome can be in free form and/or they can be part of one or more glycoconjugates. The total glycome, therefore, represents all of the carbohydrates (e.g., glycans) in the sample. The representation of the total glycome can be the number and identity of all of the carbohydrates (e.g., glycans) in the sample but is not necessarily so. The total glycome, however, does provide information regarding all of the carbohydrates (e.g., glycans) in the sample, such information can be one or more properties of the carbohydrates (e.g., glycans).

A "glycome profile" or "glycoprofile" refers to the number and kind of carbohydrates (e.g., glycans) and/or components thereof found in a sample. The glycome profile can provide, for example, the number and kind of a specific type of carbohydrate (e.g., glycan) (e.g., N-glycan, O-glycan, etc.). Each part of a glycome profile can correspond to a carbohydrate (e.g., glycan) or component thereof or a glycoconjugate or component thereof. The number refers to the amount and can be an actual or a relative amount. The "total glycome profile" or "total glycoprofile", as used herein, refers to a profile that provides information regarding one or more properties of all of the carbohydrates (e.g., glycans) in a sample. The total glycoprofile, therefore, provides the absolute or relative number and kind of all carbohydrates (e.g., glycans) and/or components thereof in a sample. The glycoprofile, in some embodiments, also provides information about carbohydrates (e.g., glycans) as part of a glycoconjugate.

To assess the glycome profile of a sample any analytical method can be used. Some of these methods are described above; others are known in the art. For example, the analytical method can be MS, NMR, HPLC, electrophoresis, capillary electrophoresis or analysis with microfluidic or nanofluidic devices. In a preferred embodiment the glycome profile is determined using a quantitative MALDI-MS or MALDI-FTMS in the presence of ATT and NafionTM coating. To quantify the glycans, in one example, calibration curves of known carbohydrate (e.g., glycan) standards can be used.

Once a glycome profile is determined, a glycome pattern can be identified. As used herein "glycome pattern" refers to a glycome profile or subset of the profile that has been associated with a certain function (of a lipid or protein), cellular state, pathological condition (i.e., a disease condition), sample, population, etc. A glycome pattern is also intended to refer to a pattern that characterizes or distinguishes a sample containing carbohydrates (e.g., glycans) from other samples. A glycome pattern can be identified using a computational

method as a glycome pattern, like the profile, can be represented by the relative or absolute amounts of components of the pattern or ratios between the components of the pattern. The glycome pattern can also be represented by combinations of different components or ratios between the components of the pattern. The pattern can also be any combination of representations, such as those provided herein. As used herein, "a component of the pattern" refers to the carbohydrates (e.g., glycans) or portions thereof that are represented by the pattern. When the carbohydrates (e.g., glycans) are part of a glycoconjugate, a component of the pattern can also be the glycoconjugate.

The pattern can be determined using a computational method. Examples of such computational methods are provided below and in the **Examples**. The computational method can, for example, incorporate one or more of the following to determine a glycome pattern: experimental data from glycome and/or carbohydrate (e.g., glycan) analysis; theoretical carbohydrate (e.g., glycan) structures; carbohydrate (e.g., glycan) composition, structure or property information from databases; carbohydrate (e.g., glycan) biosynthetic pathway information; and patient or sample origin information, such as patient history, demographics, etc. In one embodiment a method is provided whereby features from experimental data sets can be extracted and used to generate all possible data sets. The data sets can be generated from one or more analyses as provided herein and/or from databases and other tools. Such databases and tools include databases of observed carbohydrate (e.g., glycan) structures, tools to calculate mass under different conditions; tools to calculate composition, monomer (e.g., monosaccharide) content, linkage content, motif content; tools to generate theoretical structures; or some combination thereof. Databases also include data regarding patient history and related information. The method can also include submitting the combined information to a data mining analysis, establishing the relationship rules and validating the pattern. The computational method can be an iterative process. One example of a method for determining a glycome profile and pattern is provided below. Further detailed examples are provided in **Fig. 3** and in the **Examples** below.

Glycoprofiling data such as mass spectra can be generated from samples from subjects belonging to different categories. Features can be extracted from the glycoprofiling spectra. These features can be the presence or absence of one or more carbohydrates (e.g., glycans) in the profile, the relative amount of different carbohydrates (e.g., glycans) in the profile, combinations of different carbohydrates (e.g., glycans) found in the profile and/or other carbohydrate (e.g., glycan)-related properties. These carbohydrates (e.g., glycans) can be identified in the glycoprofile spectra and can be corroborated with other methods, for

instance, by using associated glycomics-based bioinformatics tools and/or a carbohydrate (e.g., glycan) databases (e.g., <http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp>).

5 The appropriate subjects can be selected for a study (e.g., based on their history in a patient database), such that the subjects chosen have the same distribution when it comes to other properties such as age, ethnicity, behavioral factors, etc. This ensures that the variation in the glycoprofiles can be attributed to the disease condition rather than other factors. The carbohydrate (e.g., glycan)-related features extracted for a population via the previous step
10 can be run through a dataset generator to create the datasets needed for pattern analysis. Different types of pattern analysis can be performed to identify the patterns in this dataset. Types of pattern analysis are known to those of ordinary skill in the art and can be found in Weiss, S. & Indurkha, N. 1998. *Predictive data mining - A practical guide*. Morgan Kaufmann, San Francisco. Three examples of patterns, rules or relationships include linear
15 discriminant, neural network and decision rules analysis.

 Once a pattern is identified using the decision set rules above, the patterns, rules or relationships can be validated. The validation can be made based on a variety of statistical methods that are used in biomarker validation as well as scientific methods to verify that the carbohydrates (e.g., glycans) found in the patterns do accurately reflect the disease state. If
20 the patterns cannot be validated, the process described above can be repeated to look for other carbohydrate (e.g., glycan)-based patterns in the glycoprofiles.

 The patterns that are ultimately validated can be recorded in a computer-generated data structure. A database of validated glycome patterns is, therefore, also provided herein.

 The patterns determined from the methods provided can provide information about a
25 sample origin, a subject's state (e.g., diseased state), etc. Patterns determined from one sample containing carbohydrates (e.g., glycans) can also be compared to patterns from other samples. Patterns that are compared can be known or unknown patterns. The patterns can represent a diseased state or a batch of glycoconjugates. Therefore, the total glycome and/or patterns deduced from the methods provided can be used for studying the effects of
30 glycosylation on protein activity and/or function as in the case of glycoprotein therapeutics. The total glycome and/or patterns deduced can also be used in methods for diagnosis, assessing prognosis and assessing drug treatment, etc.

 For example, using optimized methods described above, the total content of serum, saliva and urine glycome was analyzed, and it was shown that specific and reproducible

MALDI-MS patterns which are dependent on the source of the sample (e.g., a patient) and state (e.g., diseased state) could be obtained. Since every signal inside the pattern corresponds to specific carbohydrates (e.g., glycans), the alteration of these patterns are easily determined and correlated with the expression levels of the carbohydrates. These alterations can be easily determined manually or more efficiently with the help of computational methods. Since specific alterations in these carbohydrate (e.g., glycan) patterns can be associated with a disease state, this method serves as a reliable platform for diagnosis, prognosis or the analysis associated with therapeutics.

As stated above, the glycosylation of a protein may be indicative of a normal or a disease state. Therefore, methods are provided for diagnostic purposes based on the analysis of the total glycome or a glycome pattern. The methods provided herein can be used for the diagnosis of any disease or condition that is caused by or results in changes in glycosylation. For example, the methods provided can be used in the diagnosis of cancer, an immunological disorder, neurodegenerative disease, inflammatory disease, an infection or a genetic disorder (e.g., a congenital disorder), etc.

The diagnosis can be carried out in a subject with or thought to have a disease or condition. The diagnosis can also be carried out in a subject thought to be at risk for a disease or condition. "A subject at risk" is one that has either a genetic predisposition to have the disease or condition or is one that has been exposed to a factor that could increase the risk of developing the disease or condition.

Detection of cancers at an early stage is crucial for its efficient treatment. Despite advances in diagnostic technologies, many cases of cancer are not diagnosed and treated until the malignant cells have invaded the surrounding tissue or metastasized throughout the body. Although current diagnostic approaches have significantly contributed to the detection of cancer, they still present problems in sensitivity and specificity. Samples that are analyzed herein, therefore, can be from a subject with or be compared to a pattern associated with cancer.

Cancers or tumors include but are not limited to adrenal gland cancer, biliary tract cancer; bladder cancer, brain cancer; breast cancer; cervical cancer; choriocarcinoma; colon cancer; endometrial cancer; esophageal cancer; extrahepatic bile duct cancer; gastric cancer; head and neck cancer; intraepithelial neoplasms; kidney cancer; leukemia; lymphomas; liver cancer; lung cancer (e.g. small cell and non-small cell); melanoma; multiple myeloma; neuroblastomas; oral cancer; ovarian cancer; pancreas cancer; prostate cancer; rectal cancer;

sarcomas; skin cancer; small intestine cancer; testicular cancer; thyroid cancer; uterine cancer; urethral cancer and renal cancer, as well as other carcinomas and sarcomas.

Samples that are analyzed herein can also be from a subject with or be compared to a pattern associated with a neurodegenerative disease/disorder. "Neurodegenerative disease/disorder" is defined herein as a disorder in which progressive loss of neurons occurs either in the peripheral nervous system or in the central nervous system. As used herein "central nervous system disorders" is intended to include neurodegenerative diseases/disorders, injuries to the central nervous system (e.g., spinal cord injury), etc. Examples of neurodegenerative disorders include: (i) chronic neurodegenerative diseases such as familial and sporadic amyotrophic lateral sclerosis (FALS and ALS, respectively), familial and sporadic Parkinson's disease, Huntington's disease, familial and sporadic Alzheimer's disease, multiple sclerosis, olivopontocerebellar atrophy, multiple system atrophy, progressive supranuclear palsy, diffuse Lewy body disease, corticodentatonigral degeneration, progressive familial myoclonic epilepsy, strionigral degeneration, torsion dystonia, familial tremor, Down's Syndrome, Gilles de la Tourette syndrome, Hallervorden-Spatz disease, diabetic peripheral neuropathy, dementia pugilistica, AIDS Dementia, age related dementia, age associated memory impairment, and amyloidosis-related neurodegenerative diseases such as those caused by the prion protein (PrP) which is associated with transmissible spongiform encephalopathy (Creutzfeldt-Jakob disease, Gerstmann-Straussler-Scheinker syndrome, scrapie, and kuru), and those caused by excess cystatin C accumulation (hereditary cystatin C angiopathy); and (ii) *acute neurodegenerative disorders* such as traumatic brain injury (e.g., surgery-related brain injury), cerebral edema, peripheral nerve damage, spinal cord injury, Leigh's disease, Guillain-Barre syndrome, lysosomal storage disorders such as lipofuscinosis, Alper's disease, vertigo as result of CNS degeneration; pathologies arising with chronic alcohol or drug abuse including, for example, the degeneration of neurons in locus coeruleus and cerebellum; pathologies arising with aging including degeneration of cerebellar neurons and cortical neurons leading to cognitive and motor impairments; and pathologies arising with chronic amphetamine abuse including degeneration of basal ganglia neurons leading to motor impairments; pathological changes resulting from focal trauma such as stroke, focal ischemia, vascular insufficiency, hypoxic-ischemic encephalopathy, hyperglycemia, hypoglycemia or direct trauma; pathologies arising as a negative side-effect of therapeutic drugs and treatments (e.g., degeneration of cingulate and entorhinal cortex neurons in response to anticonvulsant doses of antagonists of the NMDA class of glutamate receptor) and Wernicke-Korsakoff's related dementia.

Neurodegenerative diseases affecting sensory neurons include Friedreich's ataxia, diabetes, peripheral neuropathy, and retinal neuronal degeneration. Neurodegenerative diseases of limbic and cortical systems include cerebral amyloidosis, Pick's atrophy, and Retts syndrome. The foregoing examples are not meant to be comprehensive but serve merely as an

5 illustration of the term "neurodegenerative disease/disorder."

Samples that are analyzed herein can also be from a subject with or be compared with a pattern associated with an immunological disorder. In one embodiment the immunological disorder is lupus. In another embodiment the immunological disorder is primary immune deficiency disease or an autoimmune disease or disorder. In yet another embodiment the

10 autoimmune disease or disorder is autoimmune deficiency syndrome (AIDS), systemic lupus erythematosus (SLE), rheumatic fever, rheumatoid arthritis, systemic sclerosis, autoimmune Addison's disease, Anklosing spondylitis or sarcoidosis.

Samples that are analyzed herein can further be from a subject with or be compared to a pattern associated with inflammation or an inflammatory disorder. In some embodiments

15 the inflammatory disorder is non-autoimmune inflammatory bowel disease, post-surgical adhesions, coronary artery disease, hepatic fibrosis, acute respiratory distress syndrome, acute inflammatory pancreatitis, endoscopic retrograde cholangiopancreatography-induced pancreatitis, burns, atherogenesis of coronary, cerebral and peripheral arteries, appendicitis, cholecystitis, diverticulitis, visceral fibrotic disorders, wound healing, skin scarring disorders

20 (keloids, hidradenitis suppurativa), granulomatous disorders (sarcoidosis, primary biliary cirrhosis), asthma, pyoderma gangrenosum, Sweet's syndrome, Behcet's disease, primary sclerosing cholangitis or an abscess. In still another embodiment the inflammatory disorder is an autoimmune condition. The autoimmune condition in some embodiments is rheumatoid arthritis, rheumatic fever, ulcerative colitis, Crohn's disease, autoimmune inflammatory

25 bowel disease, insulin-dependent diabetes mellitus, diabetes mellitus, juvenile diabetes, spontaneous autoimmune diabetes, gastritis, autoimmune atrophic gastritis, autoimmune hepatitis, thyroiditis, Hashimoto's thyroiditis, insulinitis, oophoritis, orchitis, uveitis, phacogenic uveitis, multiple sclerosis, myasthenia gravis, primary myxoedema, thyrotoxicosis, pernicious anemia, autoimmune haemolytic anemia, Addison's disease,

30 scleroderma, Goodpasture's syndrome, Guillain-Barre syndrome, Graves' disease, glomerulonephritis, psoriasis, pemphigus vulgaris, pemphigoid, sympathetic ophthalmia, idiopathic thrombocytopenic purpura, idiopathic feucopenia, Siogren's syndrome, Wegener's granulomatosis, poly/dermatomyositis or systemic lupus erythematosus.

Samples that are analyzed herein can also be from a subject with or be compared to a pattern associated with infection (e.g., *Pseudomonas* infection or *S. aureus* infection) or an infection related disorder. In some embodiments the infection is a viral infection, a bacterial infection or a fungal infection.

5 Samples that are analyzed herein can also be from a subject with or be compared to a pattern associated with a genetic disorder. As used herein, a "genetic disorder" is any disorder in which its onset or progression has a genetic basis. In some embodiments the genetic disorder is a congenital disorder, which is a condition that is genetic and is present at birth or shortly thereafter.

10 The methods provided herein also include methods for determining the glycosylation of a protein and its effects on the protein's activity and/or function. The protein glycosylation can be studied with the methods provided to determine the proper folding of the protein or to determine the influence of the protein's glycosylation on the stability and/or degradation resistance of the protein (indicative of the protein's half-life). Changing the composition or
15 the degree of glycosylation of a protein can greatly influence its half-life in circulation, as well as its activity (Chang, G.D., et al. (2003) *J Biotechnol* 102, 61-71; Perlman, S., et al. (2003) *J Clin Endocrinol Metab* 88, 3227-35.) For example, erythropoietin (EPO) is a glycoprotein that has been developed as a therapeutic due to its ability to stimulate red blood cell production in the bone marrow. It has been determined that increased sialylation of EPO
20 greatly increases its half-life in circulation (Darling, R.J., et al. (2002) *Biochemistry* 41, 14524-31.) Thus, by understanding the role of EPO glycosylation, it is possible to manufacture a more potent drug.

Similarly, methods are provided for identifying glycosylated proteins with a desired activity and/or function. For example, in immunoglobulins, glycosylation plays an important
25 role in the structure of the Fc region, which is important for activation of leukocytes expressing Fc receptors. When glycans on the IgG Fc region are truncated, the resulting conformational changes reduce the ability of the IgG to bind to the Fc receptor (Krapp, S., et al. (2003) *J Mol Biol* 325, 979-89.) In addition, IgG glycosylation is species specific, making it essential to choose the appropriate production method for protein therapeutics (Raju, T.S.,
30 et al. (2000) *Glycobiology* 10, 477-86.) For example, a human protein produced in a mouse cell line may not have the necessary carbohydrates (e.g., glycans) for optimal function in human patients. Therefore, the immune recognition of an antibody can be assessed with the methods of analysis provided herein.

Carbohydrate (e.g., glycan) patterns can also be used for determining the purity of a sample or assessing the production of a glycoconjugate.

The methods provided, where the amount or type of carbohydrates (e.g., glycans) on proteins or lipids can be determined, can be used to analyze the purity of a sample. As used herein the term “purity” refers to the proportion of a sample that contains a particular carbohydrate (e.g., glycan) or a particular glycosylation pattern. In some embodiments, the sample is determined to be at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or more pure. In some embodiments the method is used to assess the amount of a particular carbohydrate (e.g., glycan) in a sample. In some instances, it may be desired that the proteins or lipids are selected depending on the particular glycosylation pattern they exhibit. In other aspects of the invention the methods provided herein can be used to evaluate a process of producing proteins or lipids and/or compare a process with another to evaluate the types of proteins or lipids produced. The “types of proteins or lipids produced” includes not only the protein or lipid itself but also its glycosylation.

One of the major challenges during the production of glycoprotein therapeutics is to control the generation of a specific glycoform and the subsequent characterization for quality control of the product. Therefore, methods that can efficiently characterize new batches of glycoprotein therapeutics are of great value to the pharmaceutical industry. For a complete characterization of glycoprotein therapeutics, information such as glycan site occupancy, carbohydrate composition and structure at each site and quantity of each carbohydrate is required.

As described herein, the methods for analyzing a sample containing carbohydrates (e.g., glycans) can be used to assess the quality and variability of protein or lipid production. With the recently increased focus on protein-based therapeutics by pharmaceutical companies and research laboratories, it has become important to understand how glycosylation composition is influenced by production methods. In the field of bioprocess engineering, there are many different types of bioreactors available for production, e.g., protein production. Depending on the model, parameters such as pH and dissolved oxygen (DO) can be controlled in several ways, and agitation methods can result in wide variations in shear stress. In addition, the cell-feeding process during fermentation can be altered to change the cell growth profile. All of these variables can affect glycosylation—even using identical conditions in two different bioreactors causes changes in carbohydrate (e.g., glycan) patterns (Kunkel, J.P., et al. (2000) *Biotechnol Prog* 16, 462-70; Zhang, F., et al. (2002) *Biotechnol Bioeng* 77, 219-24; Senger, R.S., Karim, M.N. (2003) *Biotechnol Prog* 19, 1199-209;

Muthing, J., et al. (2003) *Biotechnol Bioeng* 83, 321-341. Therefore, provided herein are methods for analyzing the glycosylation of proteins or lipids to assess production methods and to determine the purity or homogeneity of glycosylated products produced. One example is as follows.

5 A batch of the glycoproteins can be used to generate a library of backbone-labeled peptides and glycopeptides by enzymatic digestion using methods provided herein or known in the art (Gehrmann, 2004; Yao, 2003; Reynolds, 2002; Yao, 2001). Trypsin proteolytic digest cleavage can be employed before or after carbohydrate (e.g., glycan) cleavage in order to expand the peptide library. Peptide labeling can be performed, and each characterized and
10 quantified peptide and glycopeptide can be used to generate calibration curves using LC-MS or LC-MS/MS techniques. These peptides and glycopeptides can then be mixed (in known concentrations) with the peptide/glycopeptide mixture resulting from the trypsin proteolytic cleavage digest of a sample batch under study. The co-elution of the labeled peptides with the unknown peptides followed by the co-detection (the ratio between labeled and unlabeled
15 peptides) using mass spectrometry allows the quantification of each peptide (and therefore the different glycoforms) in the sample. In addition to the peptide/glycopeptide analysis, by splitting the flow from a LC column (before entering the electrospray source) to a collection plate, the respective carbohydrates (e.g., glycans) from the eluted glycopeptides can be analyzed using the methods described herein. The use of other known methods for the
20 determination of glycan site occupancy can also be used (Cointe, 2000; Hui, 2002; An, 2003).

 The methods provided herein can also be used to determine whether or not cells are undergoing dramatic change or are "stressed cells". Stressed cells are cells that are undergoing a stress response that alters the cell's protein production. The stress response can be any change that causes altered protein production or causes the cell to deviate from its
25 normal state. Stressed cells can be identified by analyzing the carbohydrates (e.g., glycans) exhibited by the proteins on the cell's surface. Such carbohydrates (e.g., glycans) can be found in, for example, a peptide-based glycoconjugate.

 In other embodiments the methods provided are used to detect changes in glycosylation that occur under growth conditions or inflammation.

30 The samples for use in the methods provided can be any sample that contains one or more carbohydrates (e.g., glycans). The sample can be, for example, a sample of a cell, group of cells, tissue or body fluid, etc. Body fluids include serum, plasma, blood, urine, saliva, sputum, tears, CSF, seminal fluid, feces, etc. The samples can be from a subject, such as a healthy subject or diseased subject. The samples can also be from a subject undergoing a

compared to a... treatment for a disease. The sample can also be from a subject that is a healthy or non-diseased subject. Additionally, a sample can be from a pregnant woman. The sample can further be a sample of glycoconjugates, wherein the glycoconjugates are a produced therapeutic. The sample, therefore, can be a batch of glycoconjugates that have been produced.

Therefore, in other aspects of the invention methods are provided for assessing treatment regimens and/or to select specific therapies. In other aspects of the invention methods for analyzing blood type antigens are also provided.

A subject, as used herein, is any human or non-human vertebrate, e.g., dog, cat, horse, cow, pig. A sample includes any sample obtained from any of these subjects.

The present invention is further illustrated by the following **Examples**, which in no way should be construed as further limiting. The entire contents of all of the references (including literature references, issued patents, published patent applications, and co-pending patent applications) cited throughout this application are hereby expressly incorporated by reference.

EXAMPLES

Example 1 – N-Glycan Analysis

Materials and Methods

PNGase F Digest of N-glycans from Protein Cores

Between 10 and 100µg of protein were denatured for 10 minutes at 90°C with 0.5% SDS and 1% β-mercaptoethanol. Since SDS (and other ionic detergents) inhibits enzyme activity, 1% NP-40 was added to counteract these effects. The enzyme reaction was performed overnight with 2µl of PNGase F at 37°C in a 50mM sodium phosphate buffer, pH 7.5.

Purification of Released N-glycans

Proteins were precipitated with a 3X volume of 100% ethanol on ice for 1 hour. After centrifugation to remove the proteins, the supernatant containing the N-glycans was evaporated by vacuum (SpeedVac, TeleChem International, Inc., Sunnyvale, CA). Dried glycans were resuspended in 50µl of water.

Samples were desalted using 1ml ion exchange column of AG50W X-8 beads (Bio-Rad, Hercules, CA). The resin was charged with 150mM acetic acid and washed with water. Glycan samples were loaded onto the column in water and washed through with 3ml H₂O. This flow-through was collected and lyophilized to obtain the desalted sugars.

5 GlycoClean R and S cartridges were purchased from Prozyme (San Leandro, CA; formerly Glyko). GlycoClean R cartridges were primed with 3ml of 5% acetic acid, and the samples were loaded in water. Sugars were eluted with 3ml of water passed through the column. For GlycoClean S, the membrane was primed with 1ml water and 1ml 30% acetic acid, followed by 1ml acetonitrile. The glycan sample was loaded (in a maximum volume of
10 10µl) onto the disc, and the glycans were allowed to adsorb for 15 minutes. After washing the disc with 1ml of 100% acetonitrile and 5 x 1ml of 96% acetonitrile, glycans were eluted with 3 x 0.5ml water.

GlycoClean H cartridges were purchased from Prozyme (200mg bed) or ThermoHypersil (Thermo Electron Corporation, Somerset, NJ) (25mg bed). To prepare the
15 GlycoClean H cartridge, the column (containing 200mg of matrix) was washed with 3ml of 1M NaOH, 3ml H₂O, 3ml 30% acetic acid, and 3ml H₂O to remove impurities. The matrix was primed with 3ml 50% acetonitrile with 0.1% trifluoroacetic acid (TFA) (Solvent A) followed by 3ml 5% acetonitrile with 0.1% TFA (Solvent B). After loading the sample in water, the column was washed with 3ml H₂O and 3ml Solvent B. Finally, the sugars were
20 eluted using 4 x 0.5ml of Solvent A. GlycoClean H cartridges can be reused after washing with 100% acetonitrile and re-priming with 3ml of Solvent A followed by 3ml of Solvent B. For the 25mg cartridge, wash volumes were reduced to 0.5ml. Eluted fractions were lyophilized and the isolated glycans were resuspended in 10-40µl H₂O.

25 ***MALDI-MS of N-glycans***

Several MALDI-MS matrix compounds were tested in this study. First, caffeic acid was added to 30% acetonitrile to make a saturated solution, with or without 300 mM spermine. Alternatively, a saturated solution of DHB in water was used with or without 300 mM spermine. To prepare the sample spots, three methods were used. For the crushed spot
30 method, 1µl of matrix was spotted on the stainless steel MALDI-MS sample plate and allowed to dry. After crushing the spot with a glass slide, 1µl of matrix mixed 1:1 with sample was spotted on the seed crystals and allowed to dry. Alternatively, 1µl of matrix was applied followed by 1µl of sample, or vice versa. All spectra were taken with the following

instrument parameters: accelerating voltage 22000V, grid voltage 93%, guide wire 0.15% and extraction delay time of 150nsec (unless otherwise noted). All N-glycans were detected in linear mode with delayed type extraction and positive polarity.

5 Results

With an IgG-producing mouse hybridoma cell line, the effects of DO and pH control on cell metabolism and growth kinetics using two different reactor types was investigated. It was determined whether the IgG glycan profile was altered by the different reactor conditions. For a complete glycan analysis, the procedure for glycan isolation was optimized. The purification and analysis was performed using two known N-glycosylated standards with different properties, ribonuclease B (RNaseB), a glycoprotein that only contains high mannose structures (Joao, H.C., Dwek, R.A. 1993. *Eur J Biochem.* 218, 239-44), and ovalbumin, which contains both hybrid and complex glycan structures at just one glycosylation site (Harvey, D.J., et al. 2000. *J Am Soc Mass Spectrom.* 11, 564-71.) After finding the best methods for glycan analysis, the procedure was applied to samples (Hamel laboratory, MIT Bioprocess Engineering Center, Cambridge, MA) produced under various conditions.

There are several required steps for N-glycan analysis from proteins. While it is possible to study both the intact glycoprotein and glycopeptides from digested proteins, these types of analysis make it difficult to determine the exact composition of the glycan structures. Therefore, the intact glycan was removed from the core protein. Then, the sugar structures were separated from the protein core, purified and analyzed using methods that can provide specific saccharide compositions in an accurate manner.

25 ***Release and Purification of N-glycans from Protein Standards***

There are several methods, both enzymatic and chemical, to separate glycans from their protein cores. Of the chemical methods, hydrazinolysis provides the most efficient release of glycans (Patel, T., et al. 1993. *Biochemistry* 32, 679-93.) However, both N- and O-linked glycans are released using this method, and must be separated afterwards. The sample must be very clean, with no residual salts, and the reaction does not proceed efficiently in air or water, making hydrazinolysis somewhat undesirable as a quick measure of quality control.

Several enzymatic methods are available that are specific to N-linked sugars. Endo H and Endo F cleave between the two interior GlcNAc residues of the glycan core, while PNGase F cleaves between the interior GlcNAc and the asparagine side chain of the protein

core (Tarentino, A.L., et al. 1974a. *J Biol Chem.* 249, 818-24; Tarentino, A.L., Maley, F. 1974. *J Biol Chem.* 249, 811-7; Tarentino, A.L., et al. 1985. *Biochemistry* 24, 4665-71.)

Endo H only acts on high mannose or hybrid structures, while Endo F can cleave complex glycans. With Endo H and Endo F, information about fucosylation on the reducing end GlcNAc is lost since this residue remains attached to the protein core. On the other hand, PNGase F releases the entire glycans and can cleave all classes of N-glycans, making it a tool of choice, in some embodiments, for N-glycan release.

For optimal enzyme activity, proteins should be unfolded prior to digestion with PNGase F. Typically, a protein sample can be denatured by heating in the presence of β -mercaptoethanol and/or SDS. After PNGase F cleavage, samples contain a mixture of free glycans, protein, detergent (from the denaturing step) and salts. In some instances it is preferred that everything except the glycans are removed from the sample. To achieve this, the proteins were first precipitated with ethanol, and the supernatant containing the glycans was then dried under vacuum (SpeedVac) and resuspended in water. At this point, the most difficult component to get rid of was the detergent, which interferes with some types of analytical techniques.

There are several commercially available resins and cartridges for N-glycan clean-up after PNGase F digest. In addition to an ion exchange resin (AG50W X-8 from Bio-Rad), three types of purification columns from Prozyme were tested—Glyco H, S and R. Glyco R contains a reverse phase material that allows glycans to flow through, while retaining peptides and detergents. Glyco S is a small membrane that adsorbs the sugars in >90% acetonitrile, while hydrophobic molecules are washed away. The glycans can then be eluted with water. Glyco H, on the other hand, is a porous graphitic carbon matrix which retains both neutral and charged sugars, while allowing salts to be washed away with a low concentration of acetonitrile. The sugars can then be eluted with higher acetonitrile concentrations. Proteins and detergents typically remain on the Glyco H column. Overall, the Glyco H cartridge yielded the best results in these studies (Table 2).

Glycan Analysis by MALDI-MS

Numerous analytical techniques have been applied to study N-glycans, including mass spectrometry, NMR, electrophoresis, and chromatographic methods. NMR, for instance, can provide detailed structural information in a single experiment. Due to the lack of natural chromophores in N-linked carbohydrates, many of the procedures require the labeling of saccharides with chemical tags or fluorescent labels to facilitate detection. In

fluorescence assisted carbohydrate electrophoresis (FACE), glycans are fluorescently labeled and run on a polyacrylamide gel (Hu, G.F. 1995. *J Chromatogr A*. 705, 89-103.) Glycan bands can then be excised for further structural analysis. Similar methods use HPLC or CE for greater sensitivity and better separation. However, these techniques merely yield migration times of a sample's components, giving limited structural information.

One of the simplest and most sensitive glycan analysis methods is MALDI-MS, which has detection limits in the femtomole to picomole range. In addition, many samples can be analyzed in a single experiment within minutes. MALDI-MS is a soft ionization technique that utilizes an organic matrix to absorb and transfer the ionizing energy from the laser. This technique is useful for many applications, from small molecules to large proteins over 100 kDa. However, sample ionization is sensitive to instrument conditions as well as sample preparation.

In particular, the matrix used to suspend the sample is important for good ionization. The efficiency of a particular matrix can vary widely, depending on the nature of the sample. Multiple matrix preparations were tested, namely caffeic acid (saturated solution in 30% acetonitrile) with or without spermine, and DHB (saturated solution in water) with or without spermine. In addition, several spotting methods were evaluated: spotting 1 µl of sample followed by 1 µl of matrix, spotting matrix followed by sample, or mixing the two before spotting. Whether using the crushed spot method to promote matrix crystallization would improve signal intensity (Rhombert, A.J., et al. 1998. *Proc Natl Acad Sci U S A* 95, 4176-81) was also investigated. When acquiring the MALDI-MS data, the data collection was optimized by varying instrument parameters such as guide wire voltage, accelerating voltage, grid values, as well as negative vs. positive mode.

To evaluate the MALDI-MS conditions and calibrate the masses, commercially available N-glycan standards (NGA2 and NGA3) were used. In addition, RNaseB and ovalbumin were used as model glycoproteins to determine the effects of sample preparation on spectra quality and to optimize glycan release.

MALDI conditions for N-glycan analysis were optimized using the matrix and sample preparation conditions shown in **Table 2**. Among the matrix preparations used, DHB with spermine displayed the best results. Typically, spermine is used to allow glycans to be detected in negative mode, but it enhanced the glycan signals even in positive mode. The neutral glycans had poor signals in negative mode. Using the crushed spot method did not make a significant difference in signal intensity.

Geochronology

Table 2. Optimization of Conditions for MALDI-MS and N-glycan Clean-up

Sample	Matrix	Sample info	Results and comments
1. NGA2, NGA3	DHB, saturated solution in H ₂ O, 300mM spermine	1 µl matrix on plate, add 1 µl sample	Signal okay. Not too much noise.
2. NGA2, NGA3	" "	1 µl sample on plate, add 1 µl matrix	Better signal than Sample 1 or 3. This method used in all subsequent samples unless otherwise noted.
3. NGA2, NGA3	" "	Mix sample and matrix, spot 1 µl.	Lower signal than Sample 1.
4. NGA3	Caffeic acid, 30% ACN, saturated solution		Good signal intensity but significant unidentified adduct.
5. NGA3	Caffeic acid, 30% ACN, 300mM spermine		Large unidentified contamination peak.
6. NGA3	Caffeic acid, 30% ACN	Crushed spot method	Good signal intensity but also more noise than Sample 2.
7. NGA3	DHB, saturated solution in H ₂ O		Low signal intensity compared to Sample 2 or 5.
8. NGA3	DHB, saturated solution in H ₂ O	ACC voltage 18000, guide wire 0.1%.	Comparable to Sample 7.
9. NGA3	DHB, saturated solution in H ₂ O, 300mM spermine		Good signal.
10. NGA3	" "	ACC voltage 18000, guide wire 0.1%.	
11. NGA3	" "	Negative mode	Very low signal, almost undetectable.
12. 500 µg RNaseB	" "	Glyco S	Some high mannose peaks, many unidentified peaks.
13. 500 µg RNaseB	" "	AG50W X-8 column and batch mode	Both spots spread a lot, no signal.
14. 500 µg RNaseB	" "	Glyco R	Spot does not dry properly.

15. 500µg RNaseB	" "	Glyco H (200mg)	Good signal, Man-5 through Man-9.
16. 500µg Ovalbumin	" "	Glyco H (200mg)	Good signal, 30 peaks that match published reports.
17. 10µg RNaseB or ovalbumin	" "	Glyco H (25mg)	Spots spread, mostly contamination peaks in 1000-1300 Da range. Probably detergent.
18. 50µg RNaseB	" "	Glyco H (25mg)	Spot spreads a lot, significant detergent contamination peaks.
19. 15µg RNaseB	" "	Glyco H (200mg)	Good signal, very slight contamination that does not interfere with signal. Glyco H column used for future experiments.
20. 150µg RNaseB	" "	Glyco H (200mg)	Good signal, very clean.

Using commercially available glycans and known protein standards, it was determined that the optimal method for purifying glycans after PNGase F release was to use GlycoClean H cartridges containing 200mg of the stationary material. The use of this method resulted in MALDI-MS spectra of N-glycans from RNaseB and ovalbumin that were consistent with published reports. The ion exchange resin did not remove all of the detergents from the sample, causing the sample spots to spread on the MALDI-MS target and not crystallize properly. GlycoClean R, on the other hand, removed detergents but did not completely remove salt, which subsequently interfered with matrix crystallization and spectra quality. GlycoClean S yielded acceptable sample spots on the MALDI-MS target, but failed to remove all contamination.

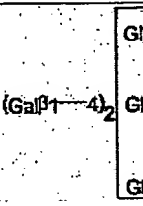
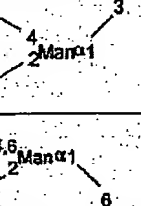
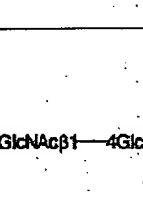
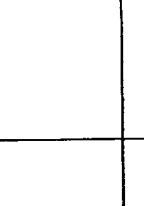


Fig. 5 shows spectra of some of the representative RNaseB samples from Table 2, with glycans purified under different conditions. In the cleanest samples, all glycan masses correspond with high mannose structures (Man-5 through Man-9).

To validate the reproducibility of the method, ovalbumin was used as a protein standard with complex type N-glycans. Optimized purification and MALDI-MS conditions were used (Glyco H 200mg, DHB matrix with spermine). The MALDI-MS data displayed results comparable to previously published reports (Harvey, D.J., et al. 2000. *J Am Soc Mass Spectrom* 11, 564-71.) Fig. 6 shows the MALDI-MS spectrum of ovalbumin glycans, and Table 3 lists the observed peaks and their structures.

Table 3. N-glycan Structures from Ovalbumin (Harvey et al, 2000)

Peak	Structure	Theoretical Mass
1		1136.4
2		1298.5
3a		1339.5
3b		1339.5
4		1460.5
5		1501.5
6a		1542.6
6b		1542.6
7		1663.6
8		1704.6

9a		1745.6
9b		1745.6
10a		1866.7
10b		1866.7
11a		1907.7
11b		1907.7
12a		1948.7
12b		1948.7

13	 <p>Chemical structure of N-glycan 13: A core fucose (Galβ1-4GlcNAcβ1-) is attached to a GlcNAcβ1-2Manα1-6GlcNAcβ1-4Manβ1-4GlcNAcβ1-4GlcNAc backbone. The branching points are at the 2 and 6 positions of the first mannose.</p>	2028.7
14a	 <p>Chemical structure of N-glycan 14a: A core fucose (Galβ1-4)2 is attached to a GlcNAcβ1-2Manα1-6GlcNAcβ1-4Manβ1-4GlcNAcβ1-4GlcNAc backbone. The branching points are at the 2 and 6 positions of the first mannose.</p>	2069.7
14b	 <p>Chemical structure of N-glycan 14b: A core fucose (Galβ1-4)2 is attached to a GlcNAcβ1-2Manα1-6GlcNAcβ1-4Manβ1-4GlcNAcβ1-4GlcNAc backbone. The branching points are at the 2 and 6 positions of the first mannose.</p>	2069.7
15a	 <p>Chemical structure of N-glycan 15a: A core fucose (Galβ1-4GlcNAcβ1-) is attached to a GlcNAcβ1-2Manα1-6GlcNAcβ1-4Manβ1-4GlcNAcβ1-4GlcNAc backbone. The branching points are at the 2 and 6 positions of the first mannose.</p>	2110.8
15b	 <p>Chemical structure of N-glycan 15b: A core fucose (Galβ1-4) is attached to a GlcNAcβ1-2Manα1-6GlcNAcβ1-4Manβ1-4GlcNAcβ1-4GlcNAc backbone. The branching points are at the 2 and 6 positions of the first mannose.</p>	2110.8
16	 <p>Chemical structure of N-glycan 16: A core fucose (GlcNAcβ1-) is attached to a GlcNAcβ1-2Manα1-6GlcNAcβ1-4Manβ1-4GlcNAcβ1-4GlcNAc backbone. The branching points are at the 2 and 6 positions of the first mannose.</p>	2151.8
17	<p>Chemical structure of N-glycan 17: A core fucose (Galβ1-4)2 is attached to a GlcNAcβ1-2Manα1-6GlcNAcβ1-4Manβ1-4GlcNAcβ1-4GlcNAc backbone. The branching points are at the 2 and 6 positions of the first mannose.</p>	2272.8

In the Applikon STR, pH can be controlled automatically by the instrument, which dispenses CO₂, NaHCO₃ and O₂ as needed. In the Wave reactor, however, measurements must be taken manually and pH adjusted by hand. The pH in this reactor can be controlled by either adding fresh media as the cells grow, or adding NaHCO₃ for increased buffering capacity, and CO₂ as needed. The main difference between the reactor types is the mode of agitation. In the Applikon STR, a blade stirrer keeps the cell suspension in motion, while a sparger introduces oxygen to the system in a controlled manner. In the Wave reactor, a rocking motion generates waves that mix the components of the system and aids the transfer of oxygen and other gases into the system.

The purified antibodies were processed according to the optimized method described above. For each sample, 100µg of protein was used as the starting material. Both positive and negative ion modes were used in the MALDI-MS to determine whether there were charged sugars present. No signal was observed in the negative mode, indicating that only neutral sugars were obtained from the antibodies. The positive ion mode MALDI-MS data of the five antibody samples are shown in Fig. 7. Glycoproteins were produced using the different conditions shown in Table 4. All fractions contained the same six glycans at 1317 Da, 1463 Da, 1478 Da, 1625 Da, 1641 Da and 1787 Da. The structures corresponding to these peaks are shown in Fig. 8 with their theoretical masses.

These results indicate that the production method did not significantly alter the occurrence of the glycans; rather, the ratios between glycans seemed to be affected. Notably, samples prepared in the Wave reactor had a lower amount of the 1625.4 Da glycan with respect to the other glycans, as well as significant reductions in the relative peak heights at 1640.9 and 1787.7 Da. Altering the culture conditions within a reactor type did not affect the relative abundance of the N-glycans.

While the exact mechanisms for producing these changes are not known, it is interesting that the largest changes occurred due to reactor type, not reactor conditions such as pH, DO or media composition. In previous studies, pH above 7.2 was shown to affect glycosylation composition (Muthing, J., et al. 2003. *Biotechnol Bioeng* 83, 321-34). However, for the two samples in this study with pH uncontrolled, the pH was between 6.8 and 7.2 throughout the culture period. Studies of DO effects on glycosylation demonstrated the largest differences at extremes (10% or 190%) (Zhang, F., et al. 2002. *Biotechnol Bioeng* 77, 219-24), while the samples studied here were produced under moderate DO conditions (between 50% and 90%). Because the Applikon STR and the Wave reactors differ most in

Table 3. their method of agitation, reactor configuration is therefore the most likely source of glycan variation.

Differences in protein glycosylation have been linked to shear stress, as can be generated by the stirring blade or the gas sparger in an STR. However, the turbulence created in the Wave reactor also generates shear stress. One hypothesis for the shear stress effect is that cells must increase their overall protein production in response to membrane and/or cytoskeletal damage. As a consequence, the biosynthetic enzymes for glycosylation are diverted away from the protein of interest (Senger, R.S., Karim, M.N. 2003. *Biotechnol Prog* 19, 1199-209).

Although most observed parameters, including total antibody production, were similar in Applikon STR and Wave cultures, cells from the Wave reactor had slight increases in metabolic rates. Changes in cell metabolism may yield effects similar to those caused by shear stress, as all glycoproteins synthesized in the cell must compete for the same machinery in the ER and Golgi.

Example 2-Profiling of N-glycans from Human Serum

Materials and Methods

Cleavage of N-glycans from Serum Glycoproteins (Reduction/Carboxymethylation Method)

Human male normal serum samples were obtained from IMPATH (Franklin, MA) and Biomedical Resources (Hatboro, PA), and stored at -85°C. For each experiment, 50µl of serum was used to harvest N-glycans. Serum samples were first diluted 1:4 with water, then DTT was added to a final concentration of 80mM. After incubation for 30 minutes at 37°C, iodoacetic acid was added to a final concentration of 400mM and incubated for 1 hour more at 37°C. The sample was dialyzed against 10mM Tris acetate pH 8.3 overnight and concentrated to ~200µl in a spin column with a 3000Da Molecular Weight Cut off (MWCO) filter (VivaScience, Hannover, Germany). To cleave the sugars from the protein, 5µl (1,000U) of PNGase F (New England Biolabs, Beverly, MA) was added and allowed to react overnight at 37°C.

Purification of N-glycans

After glycans were cleaved from the protein, the sample was dialyzed against water to remove excess salts and glycerol (from PNGase F formulation). Samples were then spun for 5 minutes at 6000xg to remove most proteins and the supernatant lyophilized to <500 μ l. A C18 cartridge (Waters Corporation, Milford, MA) was primed with 3ml methanol, then 3ml water, and 3ml 5% acetonitrile with 0.1% TFA. The supernatant from the spun down sample was applied to the cartridge, and 3ml of 5% acetonitrile with 0.1% TFA was added to elute the glycans, while unwanted proteins were retained on the column.

GlycoClean H cartridges (Prozyme) were first primed with 3ml 1M NaOH, 3ml H₂O, 3ml 30% acetic acid, and 3ml H₂O to clean the column of any impurities. Then the cartridges were washed with 3ml Solution A (50% acetonitrile, 0.1% TFA), 6ml Solution B (5% acetonitrile, 0.1% TFA), and 3ml of water. Glycan samples were loaded in a minimal volume of water (<100 μ l), and the column was washed with 3ml H₂O followed by 3ml Solution B. Neutral glycans were eluted with 3ml of 15% acetonitrile, 0.1% TFA, and acidic glycans were eluted with 3ml of Solution A. For the trials with glycan standards, the six glycans listed in Table 5 were mixed in equimolar amounts (1 μ l of 100 μ M each), and the mixture was applied to the GlycoClean H cartridge and processed as described above. Each fraction was dried, then redissolved in 40 μ l H₂O for MALDI-MS analysis. All MALDI-MS spectra were calibrated using the six glycan standards in Table 5. Separate calibration files were used for positive and negative modes.

Fractionation of Serum Proteins

Concanavalin A (ConA)-agarose beads were purchased from Vector Laboratories (Burlingame, CA). To prepare the column, 3ml ConA-agarose slurry was washed with ConA buffer (20mM Tris, 1mM MgCl₂, 1mM CaCl₂, 500mM NaCl, pH 7.4). Before loading, 500 μ l of serum was mixed with 150 μ l of 5X ConA buffer. After washing with 3ml ConA buffer, glycoproteins were eluted with 2ml of 500mM α -methyl-mannoside and dialyzed against 10mM phosphate pH 7.2 overnight at 4°C.

Protein A-agarose beads were purchased from Calbiochem (La Jolla, CA). Before use, 1ml beads were washed 3X with phosphate buffered saline (PBS). To separate IgG from other serum proteins, samples were diluted 1:4 with PBS and incubated with Protein A-agarose overnight at 4°C. Non-IgGs were collected by loading the slurry into a column and washing with 2ml PBS. IgGs were eluted with 2ml of 0.2M glycine, pH 2.5 and neutralized in 200 μ l Tris-HCl, pH 6.3.

SDS-PAGE and Glycoblotting of Serum Samples

Protein samples were prepared for SDS-PAGE by diluting 1:1 with 2X denaturing buffer (40µg/ml SDS, 20% glycerol, 30µg/ml DTT and 10µg/ml bromophenol blue in 125mM Tris, pH 6.8) and boiling for 2min. Pre-cast Nu-PAGE 10% Bis-Tris protein gels were obtained from Invitrogen (Carlsbad, CA). Each lane was loaded with a maximum of 10µl of sample, and run for 50min at 200V. After electrophoresis was complete, the gel was stained with Invitrogen SafeStain (1 hour in staining solution, then washed overnight with water).

The GlycoTrack glycoprotein detection kit was obtained from Prozyme. All reagents except buffers were supplied with the kit. Two methods were attempted—either biotinylating glycoproteins after blotting (a) or before blotting (b). For both methods, samples were first diluted 1:1 with 200mM sodium acetate buffer, pH 5.5. The membrane was blocked by incubating overnight at 4°C with blocking reagent, then washed 3x10 minutes with (Tris buffered saline (TBS).

For method (a), samples were denatured with SDS sample buffer, and subjected to SDS-PAGE and blotting to nitrocellulose. After washing the membrane with PBS, the proteins were oxidized with 10ml of 10mM sodium periodate in the dark at room temperature for 20 minutes. The membrane was washed 3 times with PBS, and 2µl of biotin-hydrazide reagent was added in 10ml of 100mM sodium acetate, 2mg/ml (ethylenediamine tetra-acetic acid) EDTA for 60 minutes at room temperature. After 3 washes with TBS, the membrane was blocked overnight at 4°C with blocking reagent. Before adding 5µl of streptavidin-alkaline phosphatase (S-AP) conjugate, the membrane was washed again with TBS. The S-AP was allowed to incubate for 60 minutes at room temperature, and excess was washed off with TBS. To develop the blot, 50µl of nitro blue tetrazolium (50mg/ml) and 37.5µl of 5-bromo-4-chloro-3-indolyl phosphate *p*-toluidine (50mg/ml) were added in 10ml TBS, 10mg/ml MgCl₂. After 60 minutes, the blot was washed with distilled water and allowed to air dry.

In method (b), 20µl of sample was mixed with 10µl of 10mM periodate in 100mM sodium acetate, 2mg/ml EDTA and incubated in the dark at room temperature for 20 minutes. To destroy excess periodate 10µl of a 12.5mg/ml sodium bisulfite solution in 200mM NaOAc, pH 5.5 was added for 5 minutes at room temperature. Biotinylation was performed by adding 5µl of biotin amidocaproyl hydrazide solution in dimethylformamide (DMF).

After incubating at room temperature for 60 minutes, the sample was mixed with SDS denaturing buffer and boiled for 2 minutes. Samples were run on SDS-PAGE gels as described above, then transferred to a nitrocellulose membrane (2hrs, 30V). At this point, blocking and developing steps were identical to method (a).

Chemical Modification of N-glycans

For permethylation, glycans in water were placed in a round-bottomed flask and lyophilized overnight. A slurry of NaOH in dimethyl sulfoxide (DMSO) (0.5ml) was added to the glycan sample, along with 0.5ml methyl iodide and incubated for 15 minutes. The sample was then diluted with water and extracted 2X with CHCl₃, collecting the organic phase. After drying the organic phase with MgSO₄, it was filtered through glass wool and dried under vacuum. Samples were then redissolved in methanol for MALDI-MS analysis.

To conjugate N-glycans to synthetic aminooxyacetyl peptide, glycans were dried and resuspended in aqueous peptide solution (240µM). After adding 1µl of 500mM NaOAc pH 5.5 and 20µl of acetonitrile, the sample was incubated overnight at 40°C. Before MALDI-MS analysis, glycopeptides were purified by C18, 0.6µl bed ZipTip (Millipore, Billerica, MA). Specifically, the tip was washed with 5µl of 100% acetonitrile, followed by water and 5% acetonitrile, 0.1% TFA. To load the sample, 2µl of sample was drawn into the tip, and discarded after 5 seconds. After washing 3X with 5µl H₂O, glycopeptides were eluted with 10% acetonitrile.

PNGase F Digestion on PVDF Membrane

PVDF-coated wells in a 96-well plate were washed with 200µl MeOH, 3x 200µl H₂O and 200µl reduction and carboxymethylation (RCM) buffer (8M urea, 360mM Tris, 3.2mM EDTA, pH 8.3). The protein samples (50µl) were then loaded in the wells along with 300µl RCM buffer. After washing the wells two times with fresh RCM buffer, 500µl of 0.1M DTT in RCM buffer was added for 1hr at 37°C. To remove the excess DTT, the wells were washed three times with H₂O. For the carboxymethylation, 500µl of 0.1M iodoacetic acid in RCM buffer was added for 30 minutes at 37°C in the dark. The wells were washed again with water, then the membrane was blocked with 1ml polyvinylpyrrolidone (360,000 average molecular weight (AMW), 1% solution in H₂O) for 1hr at room temperature. Before adding the PNGase F, the wells were washed again with water. To release the glycans, 4µl of PNGase F was added in 300µl of 50mM Tris, pH 7.5 and incubated overnight at 37°C.

Released glycans were pipetted from the wells and purified by C18 and GlycoClean H as described above.

Results

Building on the work with single protein systems, the purpose of this study was to isolate and purify N-glycans from human serum, generating a total N-glycan profile. Serum was chosen as the diagnostic medium because many disease markers are released into circulation (Pujol, J.L., et al. 2003. *Lung Cancer*. 39, 131-8; Gadducci, A., et al. 2004. *Biomed Pharmacother*. 58, 24-38), and obtaining serum is a relatively simple procedure.

Before being able to develop a method to study N-glycans from serum, it was important to understand the types of molecules present. Proteins comprise an enormous portion of serum, approximately 7% of the total wet weight (Vander, A.J., et al. 2001. *Human physiology : the mechanisms of body function*. McGraw-Hill, Boston, Mass.) Of this amount, over half is albumin (~50mg/ml), a protein that can be non-enzymatically glycosylated, but not N- or O-glycosylated (Rohovec, J., et al. 2003. *Chemistry*. 9, 2193-9.) Although the overwhelming amounts of albumin can obscure analysis for proteomics, it may not interfere with N-glycan profiling. There are also large amounts of glycosylated antibodies, which have a number of glycan structures (Bihoreau, N., et al. 1997. *J Chromatogr B Biomed Sci Appl*. 697, 123-33; Watt, G.M., et al. 2003. *Chem Biol*. 10, 807-14.) However, simple methods exist to separate these abundant antibodies from the less abundant glycoproteins.

When working with serum, there are several issues to consider that are not relevant for single protein systems. Because the proteins in the sample are so concentrated, they can easily precipitate out of solution. Also, even though albumin does not have N-linked sugars, the sheer quantity present may interfere with glycan release or purification. There are several other major proteins in serum (i.e., immunoglobulins) that are N-glycosylated, which may overshadow the signals from less abundant proteins. However, alterations in immunoglobulin glycosylation may also be correlated with changes in physiological state. To determine the contributions and/or interference from major serum proteins, several options for separating serum proteins into fractions before analysis were explored.

There are both neutral and charged sugars on serum glycoproteins. Acidic glycans generally do not ionize well in the positive ion mode of MALDI-MS, and also suffer loss of sialic acids. On the other hand, neutral sugars ionize extremely poorly in negative mode, which is commonly used for charged glycans. Therefore, a method where the neutral and

acidic structures were assayed separately was compared to two chemical modification methods that allow the glycans to ionize more uniformly.

Identifying glycan structures with complex protein mixtures can be rather difficult. By generating a master list of all possible compositions and their theoretical masses based on biosynthetic pathways for glycosylation, all possible monosaccharide composition can be assigned to each peak observed in a MALDI-MS spectrum. In most cases, each mass peak corresponds uniquely to a monosaccharide assignment. However, in some instances there can be more than one potential composition. If necessary, the correct composition can be determined by using commercially available exoenzymes that cleave the glycans only at particular linkages.

Sample Preparation

Serum samples generally contained upwards of 120mg/ml of protein, making heat denaturing less ideal. Even when diluted, the proteins in these samples precipitated rapidly, giving the sample a gel-like consistency. This could prevent the PNGase F from accessing all the N-glycan sites on the proteins. One set of samples was processed using the traditional heat-denaturation method after diluting the serum samples 1:10 in water (Fig. 9A). A number of glycan peaks were observed in the MALDI-MS spectrum, but there clearly was residual detergent contamination from the denaturing step. On a separate sample, Endo F was used since the enzyme can act on folded proteins (Fig. 9B).

However, Endo F cleaves between the first and second GlcNAc on the glycan core, causing a loss of information on core fucosylation. After Endo F digestion, the samples were purified as usual. As shown in Fig. 9, glycans could indeed be obtained using both methods. Endo F spectra had a relatively high level of baseline noise, and signal intensities were relatively low (~1000), leaving room for improvement.

As an alternative to heat denaturation, the proteins were reduced with DTT followed by carboxymethylation with iodoacetic acid to denature the proteins (Lacko, A.G., et al. 1998. *J Lipid Res.* 39, 807-20.) Reduction disrupts the disulfide bonds in proteins, while carboxymethylation prevents the proteins from re-folding. After dialysis to remove excess iodoacetic acid and DTT, PNGase F was added to the denatured proteins for overnight cleavage. An additional advantage to this method over the regular SDS/ β -mercaptoethanol heat-denaturation method was that the absence of detergents facilitated purification. After the glycans were cleaved from the core protein, the sample was dialyzed against water overnight and lyophilized. Exchanging the sample into water prepared it to be passed

through a C18 cartridge (Waters Corporation) to remove remaining protein. At this stage, both neutral and acidic sugars were present in the same sample, potentially complicating the assignment of glycan peaks in the MALDI-MS spectrum.

Because serum contains proteins with a wide variety of neutral and charged N-glycans, analysis was facilitated by separating the neutral sugars from the acidic carbohydrates. This allowed each pool to be analyzed using methods particularly suited to the chemical properties of neutral vs. charged molecules. The GlycoClean H purification cartridge (Prozyme) was used for this purpose by eluting glycan pools with different concentrations of acetonitrile. Neutral sugars were eluted with 15% acetonitrile, 0.1% TFA, while acidic sugars were eluted with 50% acetonitrile, 0.1% TFA. To test the separation of neutral and acidic sugars, six known glycan standards were used (Table 5).

Table 5. Glycan Standards Used to Test GlycoClean H Separation of Neutral and Acidic Sugars

Commercial name	Structure description	Mass	Charge state (# sialic acids)
NGA2	Asialo, agalacto biantennary	1317.2	0
NA2	Asialo, galactosylated, biantennary	1641.5	0
NA3	Asialo, galactosylated, triantennary	2006.0	0
SC1223	Disialylated, galactosylated, fucosylated biantennary	2370.2	2
A3	Trisialylated, galactosylated, triantennary	2879.9	3
SC1840	Tetrasialylated, galactosylated, tetrantennary	3683.4	4

The neutral sugars were analyzed in positive ion mode with MALDI-MS, while acidic sugars were examined in negative mode (Fig. 10). To confirm that no neutral sugars were present in the acidic glycan sample, the positive mode spectrum was checked for charged glycans. When this method was applied to human serum N-glycans, the spectra appeared much cleaner than those obtained from a mixed sample, since each group of sugars could be analyzed under optimal conditions (Fig. 11).

This process was repeated multiple times with the same serum sample to ensure reproducibility (three aliquots were purified in parallel on one day, and another two on two different days). In addition, multiple normal serum samples were processed by this method to determine the degree of glycan variation between serum samples. Five normal male human samples from each of two different sources (IMPATH tissue bank and Biomedical

Resources) were used to assess whether observed glycan profiles were consistent across suppliers. As expected, there was some variation in the spectra from different normal samples in both the neutral and acidic fraction (Fig. 11), while aliquots from the same serum sample appeared very similar even when they were purified on different days. The samples
5 from different serum banks showed similar profiles and major peak clusters.

Most Abundant Proteins

Although serum samples can be analyzed with all proteins present, including non-glycosylated species, it was determined whether better results could be obtained by removing
10 proteins such as albumin. ConA is a lectin that binds to α -linked mannose, as contained in all N-glycans (Bryce, R.A., et al. 2001. *Biophys J.* 81, 1373-88.) A serum sample was passed through a column of agarose-bound ConA. Proteins containing N-glycans bound to the column while non-glycosylated proteins were washed off (this sample was collected as the ConA flow-through). The glycoproteins were then eluted with a 500mM α -methyl-
15 mannoside solution, which competes for the ConA binding sites.

To evaluate the separation of the serum sample into glycosylated and non-glycosylated proteins, the ConA flow-through and elution samples were run on a SDS-PAGE gel (Fig. 12A). In the gel, the albumin fraction is clearly visible in the flow-through from the ConA column, while multiple bands in the elution lane represent glycoproteins. In addition,
20 the glycan profiles of both the flow-through and the elution fraction were analyzed. After dialyzing the samples against 10mM phosphate buffer, the samples were processed with PNGase F and purified by C18 cartridge and Glyco H. There were no observable glycans present in the flow-through fraction, while neutral and acidic sugars from the elution fraction are shown in Figs. 12B and 12C. The results from total serum digests, however, yield
25 MALDI-MS data with signal-to-noise ratio and signal intensity that are as good as or better than from ConA elution. Therefore, in some cases there will be little to no advantage to removing non-glycosylated proteins before analysis.

Serum samples were also depleted of antibodies through a Protein A column to determine how many major peaks in the final spectra came from IgG. The presence of
30 glycoproteins in both the flow-through and elution fractions was determined by GlycoTrack glycoprotein detection kit (Prozyme) (Fig. 13A). The Protein A elution fraction containing IgGs was treated with PNGase F and purified as described above. Figs. 13B-13E show a comparison of the glycans from IgG (Protein A elution) to the total glycan profile. Although

several of the major peaks in the spectra indeed come from this antibody population, they do not appear, at least for these samples, to be in large enough quantities to interfere with the signals from other glycans.

5 MALDI-MS Analysis of Serum N-glycans

Neutral and acidic sugars require different treatment when being analyzed by MALDI-MS. In particular, neutral sugars ionize well in the positive ion mode, but not well in negative mode, while the opposite is true for charged sugars. Three different matrix formulations were tested to determine the best one for these samples. All formulations
10 contained DHB and spermine, as this had yielded the best results with single-protein studies. The three matrix preparations were 1) saturated DHB in water with 300mM spermine, 2) 20mg/ml DHB in acetonitrile and 25mM spermine in water in a 1:1 ratio and 3) 20mg/ml DHB in methanol and 25mM spermine in water in a 1:1 ratio. Preparation 2 yielded
MALDI-MS spectra with the highest signal-to-noise ratio in both positive and negative mode,
15 and was used for all experiments.

There are several reported methods for increasing the sensitivity and ionization efficiency of mass spectrometry data in the analysis of glycans. With these methods, it is sometimes possible to analyze glycan pools as a mixture of neutral and acidic glycans, as the chemical properties of the glycans are modified to allow for more uniform ionization. Two
20 types of chemical modifications were tested to determine whether the MALDI-MS results could be improved upon.

N-glycan samples are commonly permethylated to protect each OH and NH₂ or amide group in the carbohydrate (Fukuda, M., Kobata, A. 1993. *Glycobiology : a practical approach*. IRL Press at Oxford University Press, Oxford ; New York.) This is particularly
25 useful for MS techniques such as fast atom bombardment (FAB-MS), since permethylated glycans fragment in a much more predictable manner than underivatized glycans. Permethylation can also increase sensitivity in electrospray mass spectrometry (ES-MS) and MALDI-MS. The schematic of the permethylation reaction is shown in Fig. 14. Some drawbacks to permethylation are that the sample has to be extremely clean for the reaction to
30 go to completion, and the sample requires clean-up after the reaction. In the current study, although this method slightly improved the ionization of N-glycan standards in MALDI-MS over non-modified glycans, the increase in signal-to-noise ratio was not significant (Fig. 15).

A method for increasing N-glycan ionization, as well as allowing the glycans to ionize more uniformly across species is to conjugate it to a peptide (Zhao, Y., et al. 1997. *Proc Natl*

Anal. Sci. U.S.A. 94, 1629-33.) The structure of the peptide and its glycan conjugation reaction are shown in Fig. 16. Before MALDI-MS, it was necessary to clean up the reaction mixture using a C18 ZipTip (Millipore) in order to eliminate the buffer (NaOAc) used in the reaction. The ZipTip flow-through, water wash and 10% acetonitrile elution were all spotted on the plate. The glycopeptide conjugates (in the 10% elution fraction) were readily observed with MALDI-MS, and neutral and acidic sugars ionized more evenly in the positive mode as compared to unmodified glycans (Fig. 17).

While the glycan-peptide conjugation reaction is simple, the free peptide is particularly unstable. Specifically, the peptide's active hydroxylamine group readily reacts with any aldehydes or ketones present, thus preventing it from conjugating to the glycans. Although the reaction with glycan standards displayed promising results, it was difficult to obtain a complete reaction with serum samples. Even after several attempts to label serum glycans with varying amounts of peptide, free glycan peaks in the spectra were observed from flow-through and water wash spots. Because there may be excess aldehydes or ketones remaining in serum samples, peptide conjugation was not used, and the samples were analyzed as separate neutral and acidic fractions.

Identifying Composition of Glycans from MALDI-MS Data

In a MALDI-MS spectrum, the main information obtained is mass of the parent ion. With this, it was indeed possible to deduce the monosaccharide composition of each peak (e.g., number of hexNAc, hexose, fucose and sialic acid residues). Using knowledge of biosynthetic rules, as well as whether each glycan is charged or uncharged, the number of possible structures for each mass peak observed can be significantly limited. A spreadsheet to use as a lookup table for unknown peaks was created. In addition to unmodified masses, entries for permethylated masses were included as well as peptide-conjugated glycans, according to the following equations:

	n=HexNAc	h=hexose	f=fucose	s=sialic acid
Mol. Wt.-H ₂ O=	203.1	162.1	146.1	291.3

Unmodified glycans

$$\text{mass} = 203.1n + 162.1h + 146.1f + 291.3s + 18$$

Permethylated glycans

$$\text{perm} = \text{mass} + 51 + 14[3(n+h) + 2f + 5s]$$

Peptide-conjugated glycans

$$\text{peptide} = \text{mass} + 1527.1$$

Using this table, regardless of the analytical methods, MALDI-MS peaks can be associated with specific monosaccharide compositions. Sample entries from this database are shown in Table 6.

Table 6. Table of Sample Entries for Identifying N-glycan Composition from MALDI-MS Data

HexNAc	Hexose	Fucose	Sialic Acid	mass	perm	peptide
2	3	1	0	1056.6	1345.6	2583.7
2	4	0	0	1072.6	1375.6	2599.7
3	3	0	1	1404.9	1777.9	2932
4	3	0	1	1608	2023	3135.1
4	3	2	0	1608.9	2009.9	3136
5	3	1	0	1665.9	2080.9	3193
6	3	0	0	1722.9	2151.9	3250
3	6	1	0	1746	2203	3273.1
4	3	1	1	1754.1	2197.1	3281.2
4	3	0	2	1899.3	2384.3	3426.4
4	3	2	1	1900.2	2371.2	3427.3

Using the table almost all the peaks in MALDI-MS serum profiles could be identified as glycans of known composition (Fig. 18). Many of the unidentified peaks are ammonium or sodium adducts. The composition and mass of each labeled peak are listed in Table 7. A few peaks in the acidic glycan spectrum correspond to more than one composition. This is more common in the higher mass range since there are a larger number of possible monosaccharide compositions. Many of the glycans observed in these spectra were also present in other serum samples; there are typically between 25-30 neutral glycans as well as 25-30 acidic glycans present in a given sample.

Table 7. Composition and Mass of Serum Glycans Observed in Fig. 18

Neutral glycans (Fig. 18A)					
Peak	HexNAc	Hexose	Fucose	Sialic Acid	Mass
1	2	3	1	0	1056.6
2	2	4	0	0	1072.6
3	2	5	0	0	1234.7

54	3	3	1	0	1259.7
5	3	4	0	0	1275.5
6	4	3	0	0	1316.7
7	2	6	0	0	1396.8
8	3	4	1	0	1421.8
9	3	5	0	0	1437.8
10	4	3	1	0	1642.8
11	4	4	0	0	1478.8
12	5	3	0	0	1519.8
13	2	7	0	0	1558.9
14	3	6	0	0	1599.9
15	4	4	1	0	1624.9
16	4	5	0	0	1640.9
17	5	3	1	0	1665.9
18	5	4	0	0	1681.9
19	4	5	1	0	1787.0
20	4	6	0	0	1803.0
21	5	4	1	0	1828.0
22	5	5	0	0	1844.0
23	2	9	0	0	1883.1
24	5	5	1	0	1990.1
25	5	6	0	0	2006.1
26	5	5	2	0	2136.2
27	5	6	1	0	2152.2
28	5	7	0	0	2168.2
Acidic glycans (Fig. 18B)					
Peak	HexNAc	Hexose	Fucose	Sialic Acid	Mass
1	4	4	0	1	1770.1
2	3	6	0	1	1891.2
3	4	4	1	1	1916.2
4	4	5	0	1	1932.2
5	5	4	0	1	1973.2
6	3	7	0	1	2053.3
7	4	5	1	1	2078.3
8	5	4	1	1	2119.3
9	5	5	0	1	2135.3
10	4	5	0	2	2223.5
11	4	5	2	1	2224.4
12	5	3	1	2	2248.5

13	5	4	0	2	2264.5
14	5	5	1	1	2281.4
15	5	6	0	1	2297.4
16	4	5	1	2	2369.6
17	4	5	3	1	2370.5
18	5	6	1	1	2443.5
19	5	5	1	2	2572.7
20	5	6	0	2	2588.7
21	5	6	2	1	2589.6
22	6	4	1	2	2613.7
23	5	6	1	2	2734.8
24	5	6	3	1	2735.7
25	5	6	2	2	2880.9
26	5	6	4	1	2881.8

Alternative Sample Preparation Methods

Besides performing PNGase F digests in solution, a membrane-based method was tested for the potential for high-throughput sample processing. Proteins were adsorbed onto a PVDF membrane in a 96-well plate, followed by reduction, carboxymethylation and digestion in the wells (Papac, D.I., et al. 1998. *Glycobiology*. 8, 445-54.) While this method works well with single glycoproteins, very few glycans were observed when serum was used. An explanation for this result is that albumin most likely saturated the membrane binding capacity, and most of the glycoproteins were washed away before the PNGase F was added. Without removing albumin, only a few glycans were observed in the neutral fraction, and no acidic glycans were present (Fig. 19). Although time saved by performing the experiment in a 96-well format may be negated by the extra steps required to remove albumin, using the PVDF membrane as a platform for digestion may be extremely useful for the development of a high-throughput glycomics methodology for serum samples. All samples in this study were, however, processed with the PNGase F digest in solution.

It has been demonstrated that a complete N-glycan profile from human serum proteins can be obtained. By separating glycans into neutral and acidic pools, it was possible to clearly identify glycans directly from MALDI-MS without chemical modification. In addition, it was shown that albumin and IgGs do not need to be removed from serum samples prior to analysis, although their removal can be beneficial in some contexts (e.g., in a high-throughput analysis). With the ability to profile all glycans from serum, it becomes possible to apply bioinformatics approaches to search for patterns that define normal or disease states.

Furthermore, a glycomics approach may be even more sensitive than what can be achieved with proteomics. Even in cases where protein expression does not change, the types of N-glycans present on these proteins can indicate a change in physiological condition. Already, proteomics technologies are being explored as diagnostic tools. Examining glycosylation patterns may enable more precise characterization of certain disease states, such as the differentiation between benign and malignant tumors. Thus, serum glycan profiling can advance the utility of glycomics data for the early diagnosis of currently undetectable disease states, such as, for example, in combination with a bioinformatics/computational platform.

10

Example 3-Glycan Analysis

Release of Glycans from Proteins

Several methods were used to cleave the carbohydrates from proteins:

15 A) Glycoproteins were denatured with 0.5% SDS and 1% β -mercaptoethanol. Since SDS (and other ionic detergents) inhibits enzyme activity, 1% NP-40 was added to counteract these effects. The enzymatic cleavage was performed overnight with PNGase F (New England Biolabs) at 37°C in sodium phosphate buffer, pH 7.5 or Tris acetate buffer pH 8.3.

20 B) Samples were reduced with DTT followed by alkylation with either iodoacetic acid or iodoacetamide. The sample was dialyzed against phosphate buffer, pH 7.5 or Tris acetate, pH 8.3 overnight and concentrated to ~200 μ l in a spin column with a 3000Da MWCO filter. To cleave the sugars from the protein between 100 and 2,000U of PNGase F (New England Biolabs) were used.

25 C) Glycoproteins were denatured using a buffer containing 8M urea, 3.2 mM EDTA and 360 mM Tris, pH 8.6 (Papac, D.I., et al. 1998. *Glycobiology*. 8, 445-54.) Reduction and carboxymethylation of the glycoproteins was then achieved using DTT and iodoacetic acid (or iodoacetamide), respectively. After removal of denaturing, reducing and alkylating reagents, N-glycans were selectively released from the glycoproteins by incubation with PNGase F.

30 D) The steps for protein denaturing, protein alkylation and glycan release were also performed with the proteins bound to a solid support (Papac, D.I., et al. 1998. *Glycobiology*. 8, 445-54.) PVDF-coated wells in a 96-well plate were washed with 200 μ l MeOH, 3x 200 μ l H₂O and 200 μ l RCM buffer (8M urea, 360mM Tris, 3.2mM EDTA, pH 8.3.) The protein

samples (10 to 50 μ l) were then loaded in the wells along with 300 μ l RCM buffer. After washing the wells two times with fresh RCM buffer, 300 μ l of 0.1M DTT in RCM buffer was added for 1hr at 37°C. To remove the excess DTT, the wells were washed three times with H₂O. For the carboxymethylation, 300 μ l of 0.1M iodoacetic acid in RCM buffer was added
5 for 30 minutes at 37°C in the dark. The wells were washed again with water, and the membrane was then blocked with 1ml polyvinylpyrrolidone (360,000 AMW, 1% solution in H₂O) for 1hr at room temperature. Before adding the PNGase F, the wells were washed again with water. To release the glycans, 100 to 1,000U of PNGase F were added in 300 μ l of 50mM Tris, pH 7.5 and incubated overnight at 37°C. Released glycans were pipetted from
10 the wells and purified.

E) Alternatively, after the proteins were denatured, Endo H or Endo F (instead of PNGase F) was used to release the glycans.

F) Chemical methods, such as hydrazinolysis and reductive β -elimination were also used.

15 G) The denaturing, reduction, alkylation and glycan cleavage steps were also performed in a semi-high-throughput fashion either in solution or by binding the proteins to solid supports in plates with hydrophobic membranes (Papac, D.I., et al. 1998. *Glycobiology*. 8, 445-54.)

20 ***Purification of Released N-glycans***

Several methods were used to isolate and purify the released carbohydrates. These methods were used either individually or in some combination.

A) Proteins were precipitated with a 3X volume of cold ethanol. After centrifugation to remove the proteins, the supernatant containing the N-glycans was evaporated by vacuum
25 (SpeedVac). Dried glycans were resuspended in water.

B) Concomitant protein and salt removal was achieved using cation exchange column of AG50W X-8 beads (Bio-Rad). The resin was charged with 150mM acetic acid and washed with water. Glycan samples were loaded onto the column in water, and washed through with 3ml H₂O. The flow-through was collected and lyophilized to obtain the
30 desalted sugars.

C) GlycoClean R cartridges (Prozyme) were primed with 3ml of 5% acetic acid, and the samples were loaded in water. Sugars were eluted with 3ml of water passed through the column.

D) GlycoClean S cartridges (Prozyme) were primed with 1ml water and 1ml 30% acetic acid, followed by 1ml acetonitrile. The glycan sample was loaded (in a maximum volume of 10µl) onto the disc, and the glycans were allowed to adsorb for 15 minutes. After washing the disc with 1ml of 100% acetonitrile and 5 x 1ml of 96% acetonitrile, glycans were eluted with 3 x 0.5ml water.

E) GlycoClean H cartridges (Prozyme; 200mg bed) were washed with 3ml of 1M NaOH, 3ml H₂O, 3ml 30% acetic acid, and 3ml H₂O to remove impurities. The matrix was primed with 3ml 50% acetonitrile with 0.1% TFA (Solvent A) followed by 3ml 5% acetonitrile with 0.1% TFA (Solvent B). After loading the sample in water, the column was washed with 3ml H₂O and 3ml Solvent B. Finally, the sugars were eluted using 4 x 0.5ml of Solvent A. GlycoClean H cartridges can be reused after washing with 100% acetonitrile and re-priming with 3ml of Solvent A followed by 3ml of Solvent B. For the 25mg cartridge, wash volumes were reduced to 0.5ml. Eluted fractions were lyophilized, and the isolated glycans were resuspended in 10-40µl H₂O.

F) Hypercarb SPE cartridges (Thermo Electron Corporation) were washed with 3ml of 1M NaOH, 3ml H₂O, 3ml 30% acetic acid and 3ml H₂O to remove impurities. The matrix was primed with 3ml 5% acetonitrile with 0.05% TFA (Solvent B). After loading the sample in water, the column was washed with 3ml H₂O and 3ml Solvent B. Finally, the neutral sugars were eluted using 15% acetonitrile, 0.05% TFA, and acidic glycans were eluted using 50% acetonitrile, 0.05% TFA.

G) Non-porous graphitic carbon SPE cartridges (Sigma-Aldrich, St. Louis, MO) were primed with 3ml 5% acetonitrile and 0.05% TFA (Solvent B). After loading the sample in water, the column was washed with 3ml H₂O and 3ml Solvent B. Finally, the neutral sugars were eluted using 15% acetonitrile, 0.05% TFA, and acidic glycans were eluted using 50% acetonitrile, 0.05% TFA.

H) The glycan purification step was also performed in a high-throughput format by using columns in 96-well plates. This process was facilitated by the use of a Tecan Freedom EVO automated liquid handling unit (Tecan, Durham, NC). This protocol allowed the processing of more than 90 samples at the same time.

Chemical Modification of N-glycans

Several derivatization methods can be used to increase the sensitivity and ionization efficiency in the analysis of glycans with mass spectrometry. With these methods, it is often

possible to analyze glycan pools as a mixture of neutral and acidic glycans, as the chemical properties of the glycans are modified to allow for more uniform ionization. N-glycan samples are commonly permethylated to protect each OH and NH₂ or amide group in the carbohydrate. This is particularly useful for MS techniques such as FAB-MS, since permethylated glycans fragment in a more predictable manner than underivatized glycans. Permethylation can also increase sensitivity in ES-MS and MALDI-MS.

For permethylation, glycans in water were placed in a round-bottomed flask and lyophilized overnight. A slurry of NaOH in DMSO (0.5ml) was added to the glycan sample, along with 0.5ml methyl iodide and incubated for 15 minutes. The sample was then diluted with water and extracted 2X with CHCl₃, collecting the organic phase. After drying the organic phase with MgSO₄, it was filtered through glass wool and dried under vacuum. Samples were then redissolved in methanol for MALDI-MS analysis. Some drawbacks to permethylation are that the sample has to be extremely clean for the reaction to go to completion and requires additional purification after the reaction. Although this method slightly improved the ionization of N-glycan standards in MALDI-MS over unmodified glycans, many species corresponding to incomplete modification were detected.

To conjugate N-glycans to the synthetic aminooxyacetyl peptide, glycans were dried and resuspended in aqueous peptide solution (240μM). After adding 1μl of 500mM NaOAc, pH 5.5 and 20μl of acetonitrile, the sample was incubated overnight at 40°C. Before MALDI-MS analysis, glycopeptides were purified by C18, 0.6μl bed ZipTip (Millipore). Specifically, the tip was washed with 5μl of 100% acetonitrile, followed by water and 5% acetonitrile, 0.1% TFA. To load the sample, 2μl of sample was drawn into the tip and discarded after 5 seconds. After washing 3X with 5μl H₂O, glycopeptides were eluted with 10% acetonitrile. Before MALDI-MS, it was necessary to clean up the reaction mixture using a C18 ZipTip (Millipore) in order to eliminate the buffer (NaOAc) used in the reaction. The ZipTip flow-through, water wash and 10% acetonitrile elution were all spotted on the plate. The glycopeptide conjugates (in the 10% elution fraction) were readily observed in the MALDI-MS, and neutral and acidic sugars ionized more evenly in the positive mode as compared to unmodified glycans.

While the glycan-peptide conjugation reaction is simple, the free peptide is unstable. Specifically, the peptide's hydroxylamine group readily reacts with any aldehydes or ketones present, thus preventing it from conjugating to the glycans. Other labeling reagents (i.e., 9-aminopyrene-1,4,6 trisulfonate (APTS), 9-aminonaphthalene-1,4,6 trisulfonate (ANTS), 2-

aminoacridone (AMAC), etc.) have been used, but the analysis of unmodified glycans, separated into neutral and acidic fractions, was the method used for these studies.

MALDI-MS Analysis Optimization of Unmodified Glycans

5 Neutral and acidic sugar samples can require different treatment when being analyzed by MALDI-MS. In particular, neutral sugars ionize well in the positive ion mode, while the ionization of acidic sugars is optimal using the negative ion mode. For the analysis of low abundance glycans present in a mixture of glycoforms of different glycoproteins, a matrix of matrices containing more than 96 possible recipe combinations was generated. This study
10 was designed to optimize the MALDI-MS analysis for the highest sensitivity, spot morphology, reduced peak splitting, reduced fragmentation and linear response as a function of concentration.

As a starting point, DHB was utilized in combination with spermine (20mg/ml DHB in acetonitrile and 25mM spermine in water in a 1:1 ratio.) This recipe resulted in detection
15 limits of 1pmol and 10pmol for neutral and acidic glycans, respectively. Significant peak splitting with multiple sodium and potassium ions were observed. Also, this matrix crystallized as long, needle-shaped crystals, which makes it difficult to achieve reproducible quantification of glycans present in a sample and eliminates the possibility for the automation of data acquisition.

20 Some of the matrices and reagents used in this study were: caffeic acid, DHB, spermine, 1-hydroxyisoquinoline (HIQ), ATT, 2,4,6-trihydroxyacetophenone (THAP), NafionTM, 6-hydroxypicolinic acid, 3-hydroxypicolinic, 5-methoxysalicylic acid (5-MSA), ammonium citrate, ammonium tartrate, sodium chloride, ammonium resins, etc. These reagents were used in combination with different solvents such as methanol, ethanol,
25 acetonitrile and water. The matrix of matrices study resulted in new recipes of 2,5-dihydroxybenzoic acid (5mg/ml) and 5-MSA (0.25mg/ml) in acetonitrile, for neutral glycans, and 6-aza-thiothymine (10 mg/ml in ethanol) spotted on NafionTM coating, for acidic glycans. These matrices displayed detection limits, for a mixture of carbohydrates, of 25fmol and 5fmol for neutral glycans and acidic glycans, respectively (Fig. 20). The new matrices also
30 showed minimum peak splitting, highly uniform signal intensity, spot morphology and no detectable fragmentation.

A detailed study to correlate between signal intensity, concentration and molecular weight was also performed. The analysis covered the entire range of possible molecular weights for N-glycans (approximately 900 – 4200 Da). Linear response as a function of

4 mixture of low concentration was observed for different glycans. Taken together, MALDI analysis of
glycans using these matrices can be used to quantify the amount of glycans present in a
mixture (Fig. 21). In particular, these data enable the quantification of glycans at the low
femtomole concentration range. Other methods known to those of ordinary skill in the
5 relevant art can also be used to quantify glycans at a higher range of concentrations (e.g.,
picomolar) (Harvey, D.J., *Rapid Commun Mass Spectrom.* 1993 Jul; 7(7): 614-9). For Figs.
1 and 2, the assigned peaks and labels correspond to glycan standards from Dextra
Laboratories Ltd. (Reading, United Kingdom).

A potential concern with MALDI analysis is that the ion yield of specific analyte in a
10 mixture drops as the number of constituents increase. To evaluate this, the effect of the
signal strength on the number of glycans present in a mixture was also evaluated for both
matrices. Interestingly, there was very little change in the intensity of individual glycan
signals even in the presence of other glycans, thus indicating that the ion yield of a specific
constituent is not affected by the number of analytes present in the glycan mixture (Fig. 21B).
15 This ensures that even in a complex mixture of glycans, accurate amounts can be calculated
using the signal intensity. Finally, the dynamic range of these matrices were in the low
femtomole range ensuring that changes in low abundant glycans can be accurately monitored
by using these matrices.

To prepare the sample spots, three methods were used. For the crushed spot method,
20 1 µl of matrix was spotted on the stainless steel MALDI-MS sample plate and allowed to dry.
After crushing the spot with a glass slide, 1 µl of matrix mixed 1:1 with sample was spotted
on the seed crystals and allowed to dry. Alternatively, 1 µl of matrix was applied followed by
1 µl of sample, or vice versa. When resins were used in combination with the matrices, 1 µl of
the resin was applied to the probe and allowed to dry before applying the sample in a 1:1
25 mixture with the matrix. All spectra were taken with the following instrument parameters:
accelerating voltage 22000V, grid voltage 93%, guide wire 0.15% and extraction delay time
of 150nsec (unless otherwise noted). All N-glycans were detected in linear mode with
delayed extraction and positive polarity for neutral glycans and negative polarity for acidic
glycans.

30 *LC-MS, LC-MS/MS and Capillary Electrophoresis*

Due to the limitations in isomass characterization using MALDI-MS, in some
instances other techniques such as LC-MS (or tandem-MS) and CE-LIF can be applied to

further characterize the glycans released from the glycoprotein of interest. For LC-MS (or tandem-MS), the reducing end of the carbohydrates is reduced using sodium borohydride and the carbohydrates are separated using a graphitized carbon column. The column is directly attached to an ESI-MS, which allows the detection and characterization of the carbohydrates as they elute from the column. Although the use of exoglycosidases is often added to this LC-MS analysis, MS/MS fragmentation is also used for further linkage characterization of the carbohydrates based on the fragmentation pattern.

Similarly, CE-LIF can also be used for the further separation and characterization of the glycans. In this case, the carbohydrates are first derivatized, in some embodiments, by reductive amination at their reducing end with a fluorescent molecule such as APTS, ANTS, AMAC, etc. The fluorescently-modified (or "labeled") carbohydrates are then separated by capillary electrophoresis and detected with high sensitivity via laser-induced fluorescence. Similar to LC-MS, glycosidases can also be used in combination with CE-LIF in order to get further structural linkage information on the carbohydrates.

Identifying Glycan Composition from MALDI-MS Data

In a MALDI-MS spectrum, the primary information obtained is mass of the parent ion. With this data, it was possible to deduce the monosaccharide composition of each peak (number of hexNAc, hexose, fucose and sialic acid residues). Using available information of biosynthetic rules, as well as whether each glycan is charged or uncharged, the number of possible structures for each mass peak was significantly limited. A spreadsheet to use as a lookup table for unknown peaks was created. In addition to unmodified masses, entries for permethylated masses were included as well as peptide-conjugated glycans, according to the following equations:

	n=HexNAc	h=hexose	f=fucose	s=sialic acid
Mol. Wt.-H ₂ O=	203.1	162.1	146.1	291.3

Unmodified glycans

$$\text{mass} = 203.1n + 162.1h + 146.1f + 291.3s + 18$$

Permethylated glycans

$$\text{perm} = \text{mass} + 51 + 14[3(n+h) + 2f + 5s]$$

Peptide-conjugated glycans

$$\text{peptide} = \text{mass} + 1527.1$$

Using this table, regardless of the analytical methods, mass spectrometry peaks can be associated with specific monosaccharide compositions. A table of sample entries is shown in **Table 6**. Other methods known to those of ordinary skill in the art can be used to determine the glycan identity from mass spectrometry data (See, for example, U.S. Pat. No. 5,607,859; U.S. Pat. No. 6,597,996; and WO 00/65521).

Computational Tools to Characterize Glycoprotein Mixtures

The diverse information gathered from different experimental techniques can be incorporated as constraints and used in combination with a panel of proteomics- and glycomics-based bioinformatics tools and databases for the efficient characterization (glycosylation site occupancy, quantification, glycan structure, etc.) of the glycoprotein mixture of interest (**Figs. 22 and 23**). The following six steps provides one example of how a known or unknown glycopeptide mixture can be characterized using the techniques described herein.

Step 1:

Separate the glycans from the glycopeptide mixture. Isolate and sequence the resultant peptide(s). In this example, there was only one peptide chain and that was determined to be – YCNISQKMMSRNLTKDR. This peptide has two possible N-glycosylation sites: CNIS and RNLT.

Step 2:

Digest the glycopeptide using trypsin followed by the cleavage of the glycans: one sample with ^{18}O labeling and another without labeling (^{16}O). Generate LC-MS spectra on both of the resultant samples. In this example, the following mass peaks were seen for the sample without labeling (289, 475, 476, 523, 855 and 856). With labeling the following mass peaks were seen (289, 475, 478, 523, 855, and 858).

By comparing the two spectra, the peptide fragments with mass 475 and 855 contain the glycosylation sites – both glycosylation sites are glycosylated. Based on a trypsin digest simulation of the peptide (YCNISQKMMSRNLTKDR) (See, for example, <http://us.expasy.org/tools/peptidecutter/>) the different masses were assigned as the following: 289 – DR; 475 – NLTK; 523 – MMSR; 855 – YCNISQK.

During the deglycosylation step, the Asn residue is converted to an Asp residue, which results in a total increase in molecular weight of 1Da, thus explaining the appearance of the 476 and 856 peaks. The deglycosylation with concomitant ^{18}O -labeling results in an increase of 2Da in the peptides that originally had a glycosylation site.

This explains the appearance of the 478 and 858 peaks.

The quantitative measurement of the peaks via the methods described above reveals that the glycosylation site at NLTK is 75% glycosylated. Similarly, the data for YCNISQK reveals that it is 50% glycosylated. Similarly the

undigested glycopeptide mixture is also cleaved of the glycans and label-processed as described above. The resultant analysis shows that the entire mixture is 75% glycosylated.

Step 3:

- 5 The glycans are separated and the resultant glycans analyzed through MALDI-MS. In this example, the resultant masses with relative abundance (Table 8) were:

Table 8 – Masses and Relative Abundance

Mass	Relative Abundance
1235	40
1397	44
1559	16

- 10 Thus, there are three different glycans in this glycopeptide mixture.

Step 4:

Digest the glycopeptide mixture with trypsin and analyze the resultant mixture through MALDI-MS. In this example the resultant masses are 289, 475, 523, 854, 1871, 2033 and 2089.

- 15 Based on comparing the MS results with the trypsin digest simulation of the peptide, the following observations are made. Fragment NLTK is glycosylated with glycans with a mass of 1397 or 1559. Fragment YCNISQK is glycosylated with glycans with a mass of 1235.

Thus there are six possible glycopeptide chains in the mixture.

- 20 o Chain A that is not glycosylated.
 o Chain B in which the second Asn is glycosylated with Glycan – 1397.
 o Chain C in which the second Asn is glycosylated with Glycan – 1559.
 o Chain D in which the first Asn is glycosylated with Glycan 1235.
 o Chain E in which the first Asn is glycosylated with 1235 and the second with 1397.
 25 o Chain F in which the first Asn is glycosylated with 1235 and the second with 1559.

Step 5:

Generate equations based on the experimental results and/or other data.

- 30 a,b,c,d,e and f are the relative abundances of chains A, B, C, D, E and F, respectively, and the following set of equations were generated based on the experimental results from steps 1 through 4.

$$\begin{array}{ll}
 a+b+c+d+e+f=1 & \text{6 possible chains} \\
 a+b+c=d+e+f & \text{50\% occupancy in first glycosylation site} \\
 (a+d)*3=(b+c+e+f) & \text{75\% occupancy in second glycosylation site} \\
 d+e+f=2.5*(c+f) & \text{Glycan 1235 to Glycan 1559}
 \end{array}$$

35

$$b+e = 2.75*(c+f)$$

Glycan 1397 to Glycan 1559

$$3*a = b+c+d+e+f$$

75% of glycopeptide chains are glycosylated

Solving the equations, the results are:

$$a = .25, b = .25, c = d = 0, e = .3, f = .2$$

Step 6:

The masses from step 3 can be resolved into potential glycan structures by using a glycan database lookup

(http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/searchByMw.jsp), and the exact

- 10 structure of the carbohydrates were corroborated from the glycosidase digest analysis. By putting together the results in steps 1 to 6, the unknown glycoprotein mixture was determined to be (Tables 9 and 10):

Table 9 – Glycan Identification

Peptide	Glycan Site 1	Glycan Site 2	Relative Abundance
YCN ₁ ISQKMMSRN ₂ LTkDR	None	None	.25
YCN ₁ ISQKMMSRN ₂ LTkDR	None	HEX ₆ HEXNAC ₂	.25
YCN ₁ ISQKMMSRN ₂ LTkDR	HEX ₅ HEXNAC ₂	HEX ₆ HEXNAC ₂	.3
YCN ₁ ISQKMMSRN ₂ LTkDR	HEX ₅ HEXNAC ₂	HEX ₇ HEXNAC ₂	.2

15

Table 10 – Glycan Structure

Glycan	Structure
HEX ₅ HEXNAC ₂	
HEX ₆ HEXNAC ₂	
HEX ₇ HEXNAC ₂	

Analysis of Glycosylation of Glycoprotein Standards

- As an example, the optimized procedures were performed using two known N-glycosylated protein standards with different properties, RNaseB, a glycoprotein that only contains high mannose structures, and ovalbumin, which contains both hybrid and complex glycan structures at one glycosylation site. The procedures described above were applied to
- 20

samples obtained from the Hamel laboratory (MIT Bioprocess Engineering Center, Cambridge, MA) and produced under various conditions.

Determination and Quantification of Glycosylation Site Occupancy

Before protease cleavage, the glycoproteins are first denatured in the presence of urea, reduced with DTT and carboxymethylated with iodoacetamide. To remove the denaturing reagents, the samples are concentrated using a centrifugal concentrator (3,000 MWCO) followed by buffer exchange into protease compatible buffer (50 mM ammonium bicarbonate, pH 8.5, for trypsin digest). The proteins are then cleaved by proteases followed by denaturation of proteases by boiling the sample in water and lyophilization. Glycosylation site-specific-labeling is achieved by reacting the samples with PNGase F in the presence of ^{18}O -water (Fig. 24). After desalting the glycosylated, unglycosylated and ^{18}O -labeled unglycosylated peptides through a C-18 solid phase extraction cartridge, the peptides are used in LC-MS, LC-tandem-MS, MALDI-MS, MALDI-FTMS or MALDI-TOF-TOF-MS. For this study, the unlabeled (^{16}O)- and ^{18}O -labeled samples were mixed in a 1:1 ratio before injection in order to facilitate the analysis. Other techniques for peptide sequencing can also be used at this point. The peptides were analyzed using a capillary LC-MS using a Vydac C-18 MS 5 μm (250x0.3mm) column (Grace Vydac, Hesperia, CA) coupled to a Mariner Biospectrometry Workstation (Applied Biosystems, Foster City, CA). The peptides generated from the protease cleavage were corroborated using the Swiss-Prot database (ribonuclease B, P00656, and ovalbumin, P01012).

By studying the data obtained from the differentially labeled peptides after glycan cleavage, the specific glycosylation site can be determined. The introduction of the ^{18}O at the glycosylation site is detected as a 2Da increase for a specific peptide. This data facilitate the determination of the glycosylation site and its occupancy. As determined using the peptide mass calculator from the Protein Data Bank (<http://us.expasy.org/tools/peptide-mass.html>) the tryptic digest of ribonuclease B should yield a peptide fragment with a $[\text{M}+\text{H}]^+$ of 475. 29 Da containing the glycosylation site (NLTK). Since the enzyme-mediated glycan cleavage generates an aspartic acid at the asparagine site, a peptide ion of $[\text{M}+\text{H}]^+$ of 476. 29 Da for the unlabeled peptide and a 478.29 Da for the ^{18}O -labeled peptide containing the glycosylation site was expected. As shown in Fig. 25, it was easy to identify the peptide fragment containing the glycosylation site in ribonuclease B by comparing the LC-MS data from the 1:1 mixture of $^{16}\text{O}/^{18}\text{O}$ -labeled peptides against the unlabeled sample. The presence of the +2Da species in a 1:1 ratio in the $^{16}\text{O}/^{18}\text{O}$ -labeled mixture and the absence of species

with $[M+H]^+$ of 4754.29 Da indicates that this peptide contains a glycosylation site and that it is 100% occupied in both samples of the mixture. By analyzing the differences between the peptide masses in this batch to the peptide masses from the samples not exposed to glycan cleavage, a preliminary identification of the glycans was obtained. This was further validated and quantified by analyzing the glycans separately as described below.

Release and Purification of N-glycans from Glycoprotein Standards

Several enzymatic and chemical methods were used to separate glycans from their protein cores. Of the chemical methods, hydrazinolysis provides an efficient release of glycans (Patel, T., et al. 1993. *Biochemistry*. 32, 679-93.) However, this approach requires the sample to be very clean, with no residual salts, and the reaction does not proceed efficiently in air or water, making hydrazinolysis somewhat undesirable as a quick measure of quality control. PNGase F was chosen among enzymatic methods for the cleavage of N-linked glycans since the use of other enzymes results in the loss of information, such as fucosylation at the proximal GlcNAc.

For optimal enzyme activity, proteins were unfolded, reduced and carboxymethylated prior to enzymatic digestion. Typically, the samples were denatured by heating in the presence of β -mercaptoethanol and/or SDS or by incubating at room temperature with urea, followed by reduction with DTT and carboxymethylation with iodoacetic acid or iodoacetamide. To isolate the carbohydrates from the sample, the proteins were first precipitated with ethanol, and the supernatant containing the glycans was then dried under vacuum and resuspended in water. Subsequent purification steps were required when detergents were used. Optimal results were obtained by using porous graphitic carbon columns. Neutral and charged carbohydrates were separated using these columns and eluted in mass spectrometry-compatible buffers. At this point, the most difficult component to get rid of was the detergent, which interferes with the types of analytical techniques that were used in this study.

Glycan Analysis

Different analytical techniques known in the art can be used for the glycan analysis methods. In this study, MALDI-TOF-MS was used due to its simplicity and sensitivity (e.g., low femtomole after optimizations as described herein). The MALDI-MS protocol was optimized for the detection and quantification of low abundance carbohydrates (Figs. 26 and

27). In particular, Fig. 27 shows the MALDI-MS spectrum of ovalbumin glycans. The observed peaks and their structures were found. The results are as shown above in Table 3.

RNaseB Computational Analysis

The information obtained from the previous analysis was analyzed using the computational platform that contains the proteomics- and glycomics-based bioinformatics tools and databases described herein.

The sequence of the protein backbone was determined from the proteomics database to be as follows:

MALKSLVLLS LLVLVLLLVLR VQPSLGKETA AAKFERQHMD SSTSAASSN YCNQMMKSRN₁
LTKDRCKPVN TFFVESLADV QAVCSQKNVA CKNGQTNCYQ SYSTMSITDC RETGSSSKYPN
CAYKTTQANK HIIVACEGNP YVPVHFDASV

The glycosylation site is at SNLT. It is 100 % glycosylated, and five different glycans were observed from the analysis of the glycans via MALDI-MS. The results of the computational analysis indicated that there were 5 different chains in the glycoprotein mixture as shown in Table 11 below:

Table 11-Results from the Computational Analysis

Protein Sequence	Glycan	Relative Abundance
MALKSLVLLS LLVLVLLLVLR VQPSLGKETA AAKFERQHMD SSTSAASSN YCNQMMKSRN ₁ LTKDRCKPVN TFFVESLADV QAVCSQKNVA CKNGQTNCYQ SYSTMSITDC RETGSSSKYPN CAYKTTQANK HIIVACEGNP YVPVHFDASV	HEX ₃ HEXNAC ₂	.41
MALKSLVLLS LLVLVLLLVLR VQPSLGKETA AAKFERQHMD SSTSAASSN YCNQMMKSRN ₁ LTKDRCKPVN TFFVESLADV QAVCSQKNVA CKNGQTNCYQ SYSTMSITDC RETGSSSKYPN CAYKTTQANK HIIVACEGNP YVPVHFDASV	HEX ₆ HEXNAC ₂	.29
MALKSLVLLS LLVLVLLLVLR VQPSLGKETA AAKFERQHMD SSTSAASSN YCNQMMKSRN ₁ LTKDRCKPVN TFFVESLADV QAVCSQKNVA CKNGQTNCYQ SYSTMSITDC RETGSSSKYPN CAYKTTQANK HIIVACEGNP YVPVHFDASV	HEX ₇ HEXNAC ₂	.1
MALKSLVLLS LLVLVLLLVLR VQPSLGKETA AAKFERQHMD SSTSAASSN YCNQMMKSRN ₁ LTKDRCKPVN TFFVESLADV QAVCSQKNVA CKNGQTNCYQ SYSTMSITDC RETGSSSKYPN CAYKTTQANK HIIVACEGNP YVPVHFDASV	HEX ₈ HEXNAC ₂	.14
MALKSLVLLS LLVLVLLLVLR VQPSLGKETA	HEX ₉ HEXNAC ₂	.06

AAKFERQHMD SSTSAASSSN YCNQMMKSRN₁
 LTKDRCKPVN TFFVESLADV QAVCSQKNVA
 CKNGQTNCYQ SYSTMSITDC RETGSSKYPN
 CAYKTTQANK HIIVACEGNP YVPVHFDASV

MALDI-MS Analysis of N-glycans from Antibodies Produced in Applikon and Wave Reactors

Two antibody samples produced by mouse-mouse hybridoma cells (Biokit SA) grown in an Applikon STR (Applikon Biotechnology) were analyzed, along with three samples produced in Wave reactors (Wave Biotech). The reactor conditions used are shown in Table 12.

Table 12. Reactor Conditions Used to Produce Antibody Samples

Sample	Reactor Type	DO	pH	Other
1	Applikon STR	50%	7	
2	Applikon STR	90%	Not controlled	
3	Wave	Controlled	Not controlled	
4	Wave	Controlled	7	NaHCO ₃ for pH control
5	Wave	Not controlled	7	Fresh media for pH control

In the Applikon STR reactor, pH can be controlled automatically by the instrument, which dispenses CO₂, NaHCO₃ and O₂ as needed. In the Wave reactor, however, measurements must be taken manually and pH adjusted by hand. The pH in this reactor can be controlled by either adding fresh media as the cells grow, or adding NaHCO₃ for increased buffering capacity, and CO₂ as needed. The main difference between the reactor types is the mode of agitation. In the Applikon STR, a blade stirrer keeps the cell suspension in motion, while a sparger introduces oxygen to the system in a controlled manner. In the Wave reactor, a rocking motion generates waves that mix the components of the system and aids the transfer of oxygen and other gases into the system.

The purified antibodies were processed according to the optimized method described above. For each sample, 100µg of protein was used as the starting material. Both positive and negative ion modes were used in the MALDI-MS to determine whether there were charged sugars present. No acidic glycans were observed from the analysis; which indicated neutral sugars were obtained from the antibodies. The MALDI-MS data of the five antibody

samples produced using different conditions contained the same six glycans with molecular weights of 1317 Da, 1463 Da, 1478 Da, 1625 Da, 1641 Da and 1787 Da. The corresponding structures to these glycans were determined using the methods described above and are shown in Fig. 28 with their theoretical masses.

5 These results indicate that the production method did not alter the nature of the glycans present in the samples, rather, the quantities of some glycans were affected. Notably, samples prepared in the Wave reactor displayed a 40% decrease in the 1625.4 Da glycan, as well as, a 20% reduction in the 1787.7 Da glycan with respect to samples prepared in the Applikon reactor. The other glycans remained equal.

10 While the exact mechanisms for these changes are not known, it is interesting that the largest changes occurred due to reactor type, not reactor conditions such as pH, DO or media composition. Because the Applikon STR and the Wave reactors differ most in their method of agitation, reactor configuration is therefore the most likely source of glycan variation.

15 Differences in protein glycosylation have been linked to shear stress, such as by the stirring blade or the gas sparger in an STR reactor. However, the turbulence created in the Wave reactor also generates shear stress. One hypothesis for the shear stress effect is that cells must increase their overall protein production in response to membrane and/or cytoskeletal damage. As a consequence, the biosynthetic enzymes for glycosylation are diverted away from the protein of interest (Senger, R.S., Karim, M.N. 2003. *Biotechnol Prog.* 20 19, 1199-209.)

25 Although most observed parameters, including total antibody production, were similar in Applikon STR and Wave cultures, cells from the Wave reactor had slight increases in metabolic rates. Changes in cell metabolism may yield effects similar to those caused by shear stress, as all glycoproteins synthesized in the cell must compete for the same machinery in the ER and Golgi.

Example 4-Glycome Profiling

Sample Preparation and Carbohydrate Purification

30 Samples (usually 60µl) from different body fluids (e.g., serum, saliva, urine, tears, etc.) were processed in a similar manner as described below. Although in most cases the entire glycoproteome from the sample was analyzed, in some cases, the samples were further fractionated in order to analyze a "sub-glycome" from a specific body fluid. For example, a specific subset of proteins (such as antibodies, serum albumins, and other high abundance

glycoproteins) were removed from the original serum sample in order to analyze a more specific subset of glycoproteins in more detail. For fractionation, the sample proteome was divided into "high abundance" and "low abundance" using solid supports containing antibodies, proteins and synthetic molecules specific for the desired proteins to be removed or concentrated. For example, IgGs were removed using Protein A agarose (Bio-Rad) beads, and serum albumin was removed using Affi-Blue gel (Bio-Rad). Other fractionations included the separation into acidic and basic proteome using cation and anion exchange chromatography or the separation between glycosylated and unglycosylated proteins using ConA columns. The removal of specific proteins was quantified by Western blots.

Proteins in the samples (either fractionated or unfractionated) were then denatured using a buffer containing 8M urea, 3.2mM EDTA and 360mM Tris, pH 8.6 (Papac, D.I., et al. 1998. *Glycobiology*. 8, 445-454.) Reduction and carboxymethylation of the sample proteome was then achieved using DTT and iodoacetamide, respectively. Although iodoacetic acid is often used as the alkylating agent for carboxymethylation, it is not optimal when used to analyze body fluids containing a wide range of glycoproteins, since it generally causes precipitation of most proteins. After removal of denaturing, reducing and alkylating reagents, N-glycans were selectively released from the glycoproteins by incubation with PNGase F. The steps for protein denaturing, protein alkylation and glycan release were also performed with the proteins bound to a solid support. The released carbohydrates were then purified from the proteins and separated into neutral and acidic glycans in one step using a graphitized carbon columns. The glycan purification step was also performed in a high-throughput format by using columns in 96-well plates. This process was facilitated by the use of a Tecan Freedom EVO automated liquid handling unit (Tecan). This protocol allowed for the processing of more than 90 samples at the same time.

Fractionation of Serum Proteins

As an example, to remove serum albumin and IgGs, Affi-Blue gel (Bio-Rad, 200 μ L) and Prot A (Bio-Rad, 200 μ L) were mixed in a 1:1 ratio and placed in a serum protein column. The column was washed with 1mL of compatible serum protein-binding buffer (20mM phosphate, 100mM NaCl, pH 7.2) using gravity flow. The column was placed in an empty 2mL collection tube and centrifuged at 10,000G for 20 seconds at 4°C. The flow was stopped during the sample preparation. Serum (60 μ L) was mixed with compatible serum protein-binding buffer (180 μ L), and 200 μ L of diluted serum was added to the top of the resin bed and allowed to mix with the column for 15 minutes. The column was then centrifuged at

10,000G for 20 seconds at 4°C. Using the same collection tube, the column was washed with 200µL of compatible serum protein-binding buffer and centrifuged again at 10,000G for 20 seconds at 4°C. For the removal of IgGs alone, only Protein A agarose beads were used and the binding buffer was modified to 10mM phosphate, 150mM NaCl, pH 8.2.

5 To separate glycosylated (mainly high-mannose) from unglycosylated proteins, ConA-agarose beads (Vector Laboratories) were used. To prepare the column, 3ml ConA-agarose slurry was washed with ConA buffer (20mM Tris, 1mM MgCl₂, 1mM CaCl₂, 500mM NaCl, pH 7.4). Before loading, 500µl of serum was mixed with 150µl of 5X ConA buffer. After washing with 3ml ConA buffer, glycoproteins were eluted with 2ml of 500mM
10 α-methyl-mannoside and dialyzed against 10mM phosphate, pH 7.2 overnight at 4°C.

Analysis of IgG and Serum Albumin Depletion

Samples were prepared for SDS-PAGE electrophoresis by diluting 1:1 with 2X sample buffer (120mM Tris base, 280mM SDS, 20% glycerol, 10% β-mercapto ethanol
15 (BME), 20 ng/ml bromphenol blue (BPB)), boiled for 5 minutes, and 10ul were loaded per lane in a 4-12% Bis-Tris precast gel (NPO323BOX, Invitrogen). Lane 1 contained 5ul of a standard (Precision All Blue Standard, 161-0373, Bio-Rad). The gel was run for 70 minutes at 200V. The gel was stained with SimplyBlue (LC6060, Invitrogen) according to the manufacturer. Imaging was performed on a Kodak Image Station 2000R (Kodak, Rochester,
20 NY). Another set of duplicate depleted samples were run as before. One gel was for SimplyBlue staining, and the other was transferred to a 0.20µm nitrocellulose membrane (LC2000, Invitrogen) employing an X Cell Blot Module (E19051, Invitrogen) for 70 minutes at 30V. The membrane was then blocked overnight at 4°C in 5% Blotto (sc-2325, Santa Cruz Biotechnology, Santa Cruz, CA) and then probed with 1:1000 Protein A-HRP (10-1023,
25 Zymed, San Francisco, CA) for 1 hour at 4°C and washed 4 times with washing buffer (1 x TBS: 200mM Tris base, 1.5M NaCl, pH7.5). The blot was developed with 4ml of substrate (ECL plus Western Blotting Detection System, Amersham Biosciences, Piscataway, NJ) for 2 minutes and then exposed. The bands corresponding to the treatments were manually captured as region of interest (ROI) employing the Kodak 1D Image Analysis Software
30 (Kodak), and the mean intensity was normalized to the controls.

The same blot was then washed again and re-probed with 1:1000 sheep anti-human albumin-HRP (AHP102P, Serotec, Raleigh, NC) for 1 hour at 4°C. The blot was washed again, developed and imaged (Fig. 29).

glycoproteins)

Glycoblotting of Serum Samples

Protein samples were prepared for SDS-PAGE by diluting 1:1 with 2X denaturing buffer (40µg/ml SDS, 20% glycerol, 30µg/ml DTT and 10µg/ml bromophenol blue in 125mM Tris, pH 6.8) and boiling for 2min. Pre-cast Nu-PAGE 10% Bis-Tris protein gels were obtained from Invitrogen. Each lane was loaded with a maximum of 10µl of sample and run for 50min at 200V. After electrophoresis was complete, the gel was stained with Invitrogen SafeStain (1 hour in staining solution, then washed overnight with water).

The GlycoTrack glycoprotein detection kit was obtained from Prozyme. All reagents except buffers were supplied with the kit. Two methods were attempted—either biotinylating glycoproteins after blotting (a) or before blotting (b). For both methods, samples were first diluted 1:1 with 200mM sodium acetate buffer, pH 5.5. The membrane was blocked by incubating overnight at 4°C with blocking reagent, then washed 3x10 minutes with TBS.

For method (a), samples were denatured with SDS sample buffer, subjected to SDS-PAGE and blotted to nitrocellulose. After washing the membrane with PBS, the proteins were oxidized with 10ml of 10mM sodium periodate in the dark at room temperature for 20 minutes. The membrane was washed 3 times with PBS, and 2µl of biotin-hydrazide reagent were added in 10ml of 100mM sodium acetate, 2mg/ml EDTA for 60 minutes at room temperature. After 3 washes with TBS, the membrane was blocked overnight at 4°C with blocking reagent. Before adding 5µl of S-AP conjugate, the membrane was washed again with TBS. The S-AP was allowed to incubate for 60 minutes at room temperature, and excess was washed off with TBS. To develop the blot, 50µl of nitro blue tetrazolium (50mg/ml) and 37.5µl of 5-bromo-4-chloro-3-indolyl phosphate p-toluidine (50mg/ml) were added in 10ml TBS, 10mg/ml MgCl₂. After 60 minutes, the blot was washed with distilled water and allowed to air dry.

In method (b), 20µl of sample were mixed with 10µl of 10mM periodate in 100mM sodium acetate, 2mg/ml EDTA and incubated in the dark at room temperature for 20 minutes. To destroy excess periodate 10µl of a 12.5mg/ml sodium bisulfite solution in 200mM NaOAc, pH 5.5 were added for 5 minutes at room temperature. Biotinylation was performed by adding 5µl of biotin amidocaproyl hydrazide solution in DMF. After incubating at room temperature for 60 minutes, the sample was mixed with SDS denaturing buffer and boiled for 2 minutes. Samples were run on SDS-PAGE gels and transferred to a nitrocellulose

membrane (2 hrs, 30M). After this point, blocking and developing steps were identical to method (a) (Fig. 30).

Glycan Release Using Solid Supports: PNGase F Digestion on PVDF Membrane

5 Glycans were also released using PVDF membranes as described in Papac, D.I., et al. 1998. *Glycobiology*. 8, 445-454. However, high abundance proteins were first removed before using this method, since it resulted in low recoveries when processing entire body fluids. PVDF-coated wells in a 96-well plate were washed with 200µl MeOH, 3 x 200µl H₂O and 200µl RCM buffer (8M urea, 360mM Tris, 3.2mM EDTA, pH 8.3). The protein samples
10 (50µl) were then loaded in the wells along with 300µl RCM buffer. After washing the wells two times with fresh RCM buffer, 500µl of 0.1M DTT in RCM buffer were added for 1hr at 37°C. To remove the excess DTT, the wells were washed three times with H₂O. For the carboxymethylation, 500µl of 0.1M iodoacetic acid in RCM buffer was added for 30 minutes at 37°C in the dark. The wells were washed again with water, then the membrane was
15 blocked with 1ml polyvinylpyrrolidone (360,000 AMW, 1% solution in H₂O) for 1hr at room temperature. Before adding the PNGase F, the wells were washed again with water. To release the glycans, 4µl of PNGase F were added in 300µl of 50mM Tris, pH 7.5 and incubated overnight at 37°C. Released glycans were pipetted from the wells and purified through a graphitized carbon column. Similar to protocols used for the purification of glycans
20 after performing the cleavage in solution, the purification of glycans after their release using PVDF membranes was also performed in a high-throughput format using columns in 96-well plates. This process was facilitated by the use of a Tecan Freedom EVO automated liquid handling unit (Tecan).

Glycome Analysis Using Mass Spectrometry

25 Glycan analysis was applied to the total body fluid glycome. Using the methods provided above, more than 90 samples were analyzed. Optimized MALDI-MS methods, which did not require additional labeling and purification steps and also displayed great reproducibility and sensitivity for the carbohydrate analysis, was used. As shown in Fig. 31,
30 total serum glycome profiles typically displayed between 25-30 neutral glycans as well as 25-30 acidic glycans.

Using the look-up table described previously, almost all of the peaks in MALDI-MS serum profiles could be identified as glycans of known composition. Many of the

unidentified peaks are sodium adducts. The composition and mass of each labeled peak are as listed above in Table 7. However, a few peaks in the acidic glycan spectrum correspond to more than one composition. This is more common in the higher mass range since there are a larger number of possible monosaccharide compositions.

5

Validation of Biomarker Structures

MALDI-MS analysis can be used to analyze the entire glycome profile in a sample and compare the changes in glycome composition between samples in a rapid and efficient manner. Due to the limitations in isomass characterization, in some instances, other techniques known to the art can be used to further characterize and validate the biomarkers determined from the total profile found using MALDI-MS techniques. For example, LC-MS and CE-LIF can be used in combination with a panel of exoglycosidases in order to obtain further linkage characterization of the carbohydrates (Fig. 32). After a specific pattern is established based on MALDI-MS results, and the possible species are determined, matched samples displaying the differences in patterns are analyzed by these techniques in order to come up with defined structures of the biomarkers of interest. LC-MS/MS is also used to obtain linkage information based on fragmentation patterns.

10

15

Other Body Fluids

Similar to the serum glycome analysis, the entire glycome from other body fluids such as saliva and urine have been studied (Fig. 33). For these, similar protocols to those employed for serum were used. In some instances, additional fractionation was used (e.g., if a fraction of the glycome or glycoproteome was to be studied.) The methods proved to be equally reproducible and sensitive for these other body fluids.

20

25

Glycome Analysis of Cell Surface Glycoproteins

The methods provided herein can also be applied to the glycoprofiling of cell surfaces. All cell surface glycoproteins are cleaved using methods known in the art. Briefly, to harvest glycans using protease extraction, cells are washed 3X with PBS and incubated for 20-45 minutes with trypsin/EDTA at 37°C for protease extraction. The samples are centrifuged for 10 minutes at 3000 x g to pellet the cells, and the supernatant containing glycopeptides is collected and processed using methods described herein.

30

Glycomic Pattern Analysis

The emerging field of clinical proteomics has set new avenues for the identification of potential cancer-related biomarkers. In particular, the recent introduction of proteomic pattern diagnostics (Petricoin III, E.F. et al. 2002. *Lancet*. 359, 572-577; Wulfschle, J.D., et al. 2003. *Nature Rev Cancer*. 3, 267-275; Conrads, T.P., et al. 2003. *Expert Rev Mol Diagn.* 3, 411-420) provides a promising platform for the high-throughput discovery of new and important biomarkers. Since alterations to the normal function of the glycosylation machinery have been increasingly recognized as a consistent indication of malignant transformation and tumorigenesis (Orntoft, T.F. & Vestergaard, E.M. 1999. *Electrophoresis*. 20, 362-371; Burchell, J.M., et al. 2001. *J Mam Gland Biol Neoplasia*. 6, 355-364; Brockhausen, I. 1999. *Biochim Biophys Acta*. 1473, 67-95; Dennis, J.W., et al. 1999. *Biochim Biophys Acta*. 1473, 21-34), the final glycoproteins (specifically their carbohydrate moieties) can serve as sensitive and reliable biochemical markers to numerous diseases including cancer.

Methods for glycomic pattern analysis where the total profile of carbohydrates from body fluids or tissues can be examined in a rapid format are provided herein. This approach provides an efficient overview of the total changes in carbohydrate composition of a tissue or body fluid as a result of pathological alterations and should be reliable in sensing susceptible physiological changes to the body's natural homeostasis. The methods not only serve as fast diagnostic/prognostic tools but can also help to understand the function of specific carbohydrate modifications in some diseases. The methods also provides a reliable system to efficiently monitor the effects of therapeutics.

For instance, the optimization of MALDI-MS analysis allows reliable reproducibility that enables the fast evaluation of alterations to glycomic patterns and their subsequent association to pathological/physiological changes to a sample donor. The optimized detection limits for this method (low femtomole) allows the detection of low abundance species associated with diseases. Every signal in the pattern is rapidly correlated to the glycan identity and can be further validated using a panel of glycosidases and/or other techniques. This prevents erroneous identification, as has sometimes been the case in the field of proteomic pattern diagnostics. The pattern alterations can be easily determined manually or more efficiently with the aid of bioinformatics tools. In some cases the decreasing levels of circulating glycoproteins in serum are easily matched to the analyzed glycans. As shown in Fig. 34, the glycome profile from serum with low IgG levels, reflects the specific decrease in the respective IgG glycans with molecular weights of 1463, 1626,

1666, 1788, 1829, 2102, and 1844. These glycans have been previously shown to be attached to IgG molecules in serum (Butler, M. et al. 2003. *Glycobiology*. 13, 601-622.)

By applying a "glycomic pattern diagnostics" platform to different body fluids from patients with well-defined demographics, specific alterations in the glycomic pattern that can be correlated to the pathological state of the donor can be determined. For example, glycomic patterns have now been associated to prostate cancer by studying the serum from prostate cancer patients (Fig. 35). Glycomic patterns from the saliva of patients with viral infections have also been established. Since every signal inside the pattern corresponds to specific glycans, the alteration of these patterns are easily determined and correlated with the expression levels of the carbohydrates, such as with the methods provided herein. The specific alterations in these glycan patterns are associated with a disease state. Therefore, the methods provided serve as a reliable platform for diagnosis, prognosis and monitoring the effects of therapeutics.

Computational Pattern Analysis of Glycoprofile

The following is an example of a computational approach for identifying glycan-based biomarkers for specific diseases using data from glycoprofilng. The different steps of the process are illustrated in Fig. 36.

Using prostate cancer as an example, the goal is to identify glycan-based markers for individuals with prostate cancer. In this example, there are three possible categories of individuals – individuals with prostate cancer, benign prostatic hyperplasia (BPH) and individuals that are normal (i.e., they have a healthy, non-diseased prostate.)

Glycoprofilng data such as mass spectra are generated from samples from patients belonging to the different categories. Features are extracted from the glycoprofilng spectra. These features can be the presence or absence of one or more glycans in the profile, the relative amount of different glycans in the profile, combinations of different glycans found in the profile and/or other glycan-related properties. These glycans are identified in the glycoprofile spectra and can be corroborated with other methods, for instance, by using associated glycomics-based bioinformatics tools and/or a glycan database (<http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp>).

The appropriate patient population is selected for the study (based on their history in a patient database), such that the subjects chosen in the different categories of prostate cancer,

BPH and normal have the same distribution when it comes to other properties such as age, ethnicity, behavioral factors, etc. This ensures that the variation in the glycan profiles can be attributed to the disease condition rather than other factors. The glycan related features extracted for this population via the previous step is run through a dataset generator to create the datasets needed for pattern analysis. Different types of pattern analysis are performed to identify the patterns in this dataset. Types of pattern analysis are known to those of ordinary skill in the art and can be found in Weiss, S. & Indurkha, N. 1998. *Predictive data mining - A practical guide*. Morgan Kaufmann, San Francisco. Three examples of patterns, rules or relationships that can be identified are as follows:

- *Linear Discriminant*: The pattern identified is in the form of weights (w_{11} , w_{12} , etc.) for the different glycan related features (G_1 , G_2 , etc.) as they are related to property or class of interest (prostate cancer, BPH or normal).

- $w_{11}G_1 + w_{12}G_2 + \dots + w_{1m}G_m + w_1 = \text{prostate cancer}$
- $w_{21}G_1 + w_{22}G_2 + \dots + w_{2m}G_m + w_2 = \text{BPH}$

- *Neural Network*: The neural network identifies non linear relationships or patterns between the different features and the property or class of interest.

- $\text{net}_j = \sum W_{ij} * f_i + C_j$
- $d_j = 1 / (1 + e^{-\text{net}_j})$, where d_j can be prostate cancer, BPH or normal

- *Decision Rules*: The pattern identified is in the form of IF-THEN rules, for example

(IF G_{11} is present and G_7 is not present) or (IF G_8 is present and G_9 is present) THEN Class = prostate cancer

(IF G_1 is present and G_2 is present and G_3 is not present) THEN Class = BPH

Otherwise Class = normal

Once a pattern is identified using the decision set rules above, the patterns, rules or relationships are validated. The validation can be made based on a variety of statistical methods that are used in biomarker validation as well as scientific methods to verify that the glycans found in the patterns do accurately reflect the disease state. If the patterns cannot be validated, the process described above can be repeated to look for other glycan-based patterns in the glycoprofiles.

Use of Human Glycome for the Profiling of Populations: Population-tailored Treatments

It is documented that different people react differently to certain drugs. In many instances this might be a result of drug interference with other inherent molecular components. Also, the down-regulation of enzymes may prevent the metabolism of some drugs or their byproducts. The recent emphasis in the development of new carbohydrate-

based therapeutics will face major challenges (in comparison to protein-based drugs) due to limited availability of glycomic information and understanding in the field.

More information of all human molecular components will significantly facilitate the design and prescription of medications to specific populations (personalized medicine). The efficient analysis of the entire glycome from body fluids not only serves as a reliable diagnosis/prognosis platform but can be valuable for the profiling of populations. The generation of a human glycome databank from different ethnic groups, gender, ages, diseases, etc. will become of enormous value for current and future development and applications of drugs that might interfere with carbohydrate function. For example, the overexpression of some carbohydrates in a specific population will aid in the design and prescription of therapeutics that might interact with these carbohydrates. This information will also aid in prospective studies for the selection of dosing, activity monitoring and efficacy endpoints.

Example 5 – Use of Optimized MALDI-MS Conditions for the Improved Analysis of Highly Acidic Polysaccharides

Mass spectrometry has been used as a major tool in the analysis of highly acidic polysaccharides, such as GAGs. MALDI-MS, in particular, has been a key component in the characterization of these biopolymers. However, major experimental disadvantages still exist with the current methods. Due the complex nature of these polysaccharides, MALDI-MS characterization usually reveals multiple species. Mass spectra are usually complicated by the multiple ions complexed with the biopolymers due to their highly acidic nature. Therefore, the multiple peaks arising from the different carbohydrate-ions complexes hamper the correct assignment of the polysaccharide identity. Additionally, the splitting of one species into multiple ionic complexes decreases the effective concentration of each species into all the possible complexes resulting in decreased sensitivity. In order to test the improved matrix conditions on a highly acidic polysaccharide, hyaluronic acid was digested with hyaluronidase, fractionated via size exclusion and anion exchange chromatography, and the fragments were analyzed using MALDI-TOF-MS. Applying the optimized conditions for MALDI-MS analysis of the HA fragments, eliminated the multiple carbohydrate-ion complexes and significantly increased the sensitivity of the method (Fig. 37).

Example 6 - Computational Method Combining NMR Spectroscopy and MALDI-MS for Characterization of Glycans in a Mixture

MALDI-MS analysis of a mixture provides exact molecular weights of the glycans in the mixture. Each mass peak corresponds to a single or multiple unique monosaccharide compositions in terms of hexoses, HexNAcs, fucoses, sialic acids, etc. Each of these compositions in turn correspond to a single or multiple explicit monosaccharide compositions, such as Glc, Gal, Man, GlcNAc, GalNAc, Fuc, NeuAc and NeuGc, etc.

While MALDI-MS is not fully quantitative, the methodology has been optimized to provide a reasonably accurate quantification of the relative amounts of each of the mass peaks. Incorporating NMR data as constraints to further refine information from MALDI-MS enables the elimination of explicit compositions that do not satisfy the monosaccharide composition data from NMR and a more quantitative determination of the abundance of monosaccharides and linkage distributions. In addition, biosynthetic rules and database look-ups (e.g.,

<http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp>) can help in further convergence of the solution to obtain an accurate picture of the

number and relative abundance of the species in the sample as well as the best characterization of the individual structures corresponding to these species. A schematic of an example of this methodology is provided in Fig. 38. In this example, the starting sample was prepared by mixing 3 N-glycan standards in different proportions to obtain the specific relative abundance. As described above, MALDI-MS analysis was performed by mixing the glycan sample with the 5-MSA/DHB matrix for positive polarity and with ATT (on a NafionTM-coated plate) for negative polarity. For NMR experiments, the glycan mixture was dissolved in D2O and lyophilized 3 times. The sample was dissolved in 500 uL of D2O and 2D-COSY and 1-D 1H-spectra were recorded using Bruker 600 MHz NMR (Massachusetts Institute of Technology NMR Facility, National Institutes of Health Grant 1S10RR133886-01) using sodium 3-(trimethylsilyl)propionate-2,2,3,3-d₄ (TMSP) as internal standard.

MALDI-MS Data

Mass	Relative Intensity
1990	75%
2047	25%

Monosaccharide Composition Obtained from NMR Data

Monosaccharide	Relative Abundance
-----------------------	---------------------------

5	GlcNAc	40.9%
	Man	27.3%
	Gal	18.2%
	Fuc	6.8%
5	GalNAc	6.8%

Linkage Abundance Obtained from NMR Data

	Linkage	Relative Intensity
	Man α 6Man	10%
10	Man β 4GlcNAc	10%
	GlcNAc β 4GlcNAc	10%
	Man α 3Man	10%
	GlcNAc β 6Man	2.5%
	GlcNAc β 4Man	2.5%
15	GlcNAc β 2Man	20%
	Gal β 4GlcNAc	15%
	Gal α 3Gal	5%
	Fuc α 6GlcNAc	7.5%
20	GalNAc β 4GlcNAc	7.5%

Steps Involved in the Computational Method

1. From the masses, possible compositions were obtained:
 - a. Mol. Wt. 1990: Hex3Fuc2HexNAc3NeuAc2 or Hex5Fuc1HexNAc5 but Hex3Fuc2HexNAc3NeuAc2 is not possible because of the monosaccharide data and negative polarity MALDI-MS.
 - b. Mol. Wt. 2046: Hex5HexNAc6 or Hex10HexNAc2
2. From the composition, NMR monosaccharide information and biosynthetic rules and structures found in a carbohydrate data bank, the following glycans are possible:
 - a. Hex5Fuc1HexNAc5 : Man3Gal2Fuc1GlcNAc4GalNAc1, Man3Gal2Fuc1GlcNAc5
 - b. Hex5HexNAc6: Man3Gal2GlcNAc5GalNAc1, Man3Gal2GlcNAc6, Man5GlcNAc6
 - c. Hex10HexNAc2 : Man10GlcNAc2
3. Let,
 - a = the relative amount of Man3Gal2Fuc1GlcNAc4GalNAc1
 - b = the relative amount of Man3Gal2Fuc1GlcNAc5
 - c = the relative amount of Man3Gal2GlcNAc5GalNAc1
 - d = the relative amount of Man3Gal2GlcNAc6
 - e = the relative amount of Man5GlcNAc6
 - f = the relative amount of Man10GlcNAc2

Equations to match relative abundance information from NMR and MALDI

1. $a+b = 3*(c+d+e+f)$ -from MALDI
2. $(3/11)*(a+b+c+d) + (5/11)*e + (10/12)*f = .273$ -from Man composition (NMR)
- 45 3. $(2/11)*(a+b+c+d) = .182$ -from Gal composition (NMR)
4. $(1/11)*(a+b) = .068$ -from Fuc composition (NMR)

5. $(4/11)*a+(5/11)*(b+c)+(6/11)*(d+e)+(2/11)*f = .409$ -from GalNAc composition (NMR)
6. $(1/11)*(a+c) = .068$

Solving the set of 6 equations leads to the following result:

- 5 $a=50\%$, $b = 25\%$, $c = 25\%$, $d=e=f=0$.

Final Convergence Based on Linkages from NMR, Biosynthesis Rules and Database Look-up

- Of the following explicit compositions, Man3Gal2Fuc1GlcNAc4GalNAc1,
10 Man3Gal2Fuc1GlcNAc5 and Man3Gal2GlcNAc5GalNAc1, structures that contain only the
links in the NMR linkage table and satisfy biosynthetic rules are shown in Figs. 39A-C.
These results show an exact correlation with the initial composition of the sample.

- Each of the foregoing patents, patent applications and references that are recited in
15 this application are herein incorporated in their entirety by reference. Having described the
presently preferred embodiments, and in accordance with the present invention, it is believed
that other modifications, variations and changes will be suggested to those skilled in the art in
view of the teachings set forth herein. It is, therefore, to be understood that all such
variations, modifications, and changes are believed to fall within the scope of the present
20 invention as defined by the appended claims.

We claim:

CLAIMS

1. A method of analyzing a sample containing one or more glycoconjugates, which comprise one or more glycans conjugated to a non-saccharide component, comprising:
 - 5 (a) analyzing the glycoconjugates to characterize the glycoconjugates,
 - (b) analyzing the non-saccharide components of the glycoconjugates to characterize the non-saccharide components,
 - (c) separating the glycans from the sample containing one or more glycoconjugates,
 - (d) analyzing the glycans to characterize the glycans, and
 - 10 (e) determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample with the results obtained from steps (a), (b) and (d) and a computational method.
2. The method of claim 1, wherein the computational method comprises generating
15 constraints from the results obtained from steps (a), (b) and/or (d) and solving them.
3. The method of claim 2, wherein further constraints are generated from databases containing information about glycans, non-saccharide components, glycoconjugates or a combination thereof.
20
4. The method of claim 2, wherein further constraints are generated from biosynthetic rules.
5. The method of claim 2, wherein further constraints are generated from information about the sample origin.
25
6. The method of claim 5, wherein the information about the sample origin comprises information regarding the expression system or expression conditions for the synthesis of the glycoconjugates, the species from which the glycoconjugates are derived, the expression levels of glycosidases and glycosyltransferases from the source from which the
30 glycoconjugates are obtained or the state of the source from which the glycoconjugates are obtained.

7. The method of any of claims 2-6, wherein the constraints are one or more mathematical equations and are solved by determining the solution of the one or more mathematical equations.

5 8. The method of claim 7, wherein the constraints are solved with a computer program.

9. The method of claim 1, wherein nuclear magnetic resonance (NMR) is used to analyze the glycoconjugates.

10 10. The method of claim 1, wherein the glycosylation sites and glycosylation site occupancy of the glycoconjugates are determined.

11. The method of claim 1, wherein the determination of the glycosylation sites and glycosylation site occupancy, comprises:

15 cleaving the non-saccharide components of the glycoconjugates,
 cleaving the glycans from the non-saccharide components and labeling the non-saccharide components of a first portion of the sample at the glycosylation sites,
 cleaving the glycans from the non-saccharide components of a second portion of the sample,
20 analyzing the first and second portions of the sample containing the non-saccharide components, and
 comparing the results.

12. The method of claim 11, wherein the non-saccharide components of the first portion of
25 the sample are labeled with a labeling agent.

13. The method of claim 12, wherein the labeling agent is an isotope of C, N, H, S or O.

14. The method of claim 13, wherein the labeling agent is ^{18}O .

30

15. The method of claim 13, wherein the labeling agent is ^2H .

16. The method of claim 11, wherein the non-saccharide components of the second portion are unlabeled.

from GalNAc

from GalNAc

17. The method of claim 11, wherein the non-saccharide components of the second portion are labeled.
- 5 18. The method of claim 11, wherein the glycosylation site occupancy is quantified from ratios of the masses of the non-saccharide components of the first and second portions of the sample.
19. The method of claim 11, wherein the first and second portions of the sample are analyzed
10 with a mass spectrometric method.
20. The method of claim 19, wherein the mass spectrometric method is LC-MS, LC-MS/-MS, MALDI-MS, MALDI-TOF-MS, MALDI-TOF PSD-MS, MALDI-TOF/TOF-MS, MALDI-TOF/TOF-MS/MS, MALDI-TOF/TOF PSD-MS, MALDI-FTMS, LC-MALDI-
15 TOF/TOF-MS, Nano-LC MALDI-TOF/TOF-MS, Nano-LC MALDI-TOF/TOF PSD-MS, Nano-LC MALDI-TOF/TOF-MS/MS or TANDEM-MS.
21. The method of claim 11, wherein the first and second portions of the sample are analyzed separately.
- 20 22. The method of claim 11, wherein the first and second portions of the sample are analyzed as a mixture.
23. The method of claim 1, wherein step (b) comprises determining the sequence of the non-
25 saccharide components.
24. The method of claim 23, wherein the non-saccharide components are peptides and a peptide sequence is determined.
- 30 25. The method of claim 23, wherein the sequence of the non-saccharide components is determined prior to or subsequent to the analysis of the glycoconjugates.

26. The method of claim 1, wherein the glycans are analyzed with a mass spectrometric method, an electrophoretic method, NMR, a chromatographic method or a combination thereof.

27. The method of claim 26, wherein the mass spectrometric method is ESI-MS, LC-MS, LC-MS/MS, MALDI-MS, MALDI-MS/MS, MALDI-TOF-MS, MALDI-TOF PSD-MS, MALDI-TOF/TOF-MS, MALDI-TOF/TOF-MS/MS, MALDI-TOF/TOF PSD-MS, MALDI-FTMS, LC-MALDI-TOF/TOF-MS, Nano-LC MALDI-TOF/TOF-MS, Nano-LC MALDI-TOF/TOF PSD-MS, Nano-LC MALDI-TOF/TOF-MS/MS or TANDEM-MS.

28. The method of claim 26, wherein the electrophoretic method is capillary electrophoresis (CE) or CE-LIF.

29. The method of claim 26, wherein the chromatographic method is HPLC.

30. The method of claim 26, wherein the glycans are analyzed with NMR.

31. The method of claim 30, wherein the results from the NMR provide monosaccharide composition and linkage information for the glycans.

32. The method of claim 30, wherein the method further comprises analyzing the glycans with MALDI-MS.

33. The method of claim 32, wherein the results from the MALDI-MS provide monosaccharide composition and relative abundance information for the glycans.

34. The method of any of claims 26 and 30-33, wherein the mass spectrometric method is performed in the presence of a thymine derivative and an ion exchange resin.

35. The method of claim 34, wherein the thymine derivative is 6-aza-2-thiothymine (ATT).

36. The method of claim 34, wherein the ion exchange resin is a perfluorinated ion exchange resin.

37. The method of claim 36, wherein the perfluorinated ion exchange resin is Nafion™.

38. The method of any of claims 26 and 30-33, wherein the method further comprises contacting the glycans with one or more glycan-degrading enzymes.

5

39. The method of claim 38, wherein the one or more glycan-degrading enzymes is sialidase, galactosidase, mannosidase, N-acetylhexosaminidase or a combination thereof.

40. The method of any of claims 26 and 30-33, wherein the method further comprises
10 contacting the glycans with strong acidic or basic conditions.

41. The method of any of claims 1, 26 and 30-33, wherein the step of analyzing the glycans comprises quantifying the glycans using calibration curves of known glycan standards.

42. The method of any of claims 1, 26 and 30-33, wherein the method further comprises
15 purifying the glycans.

43. The method of claim 42, wherein the glycans are purified with solid phase extraction cartridges or ion exchange resins.

20

44. The method of claim 43, wherein the solid phase extraction cartridges are graphitic carbon columns, non-graphitic carbon columns or C-18 columns.

45. The method of any of claims 1, 26 and 30-33, wherein step (c) comprises cleaving with
25 PNGase F, Endo H, Endo F, hydrazinolysis or alkaline borohydride.

46. The method of any of claims 1, 26 and 30-33, wherein step (c) comprises denaturing the glycoconjugates with a denaturing agent.

47. The method of claim 46, wherein the denaturing agent comprises a detergent, urea, high
30 salt concentration, guanidium hydrochloride or heat.

48. The method of claim 46, wherein the glycoconjugates are reduced following their denaturation with a reducing agent.

The method of c)

49. The method of claim 48, wherein the reducing agent comprises dithiothreitol (DTT), β -mercaptoethanol or Tris(2-carboxyethyl)phosphine (TCEP).

5 50. The method of claim 48, wherein the glycoconjugates are alkylated with an alkylating agent following their reduction.

51. The method of claim 50, wherein the alkylating agent is iodoacetic acid or iodoacetamide.

10

52. The method of any of claims 1, 26 and 30-33, wherein the glycans are unmodified.

53. The method of any of claims 1, 26 and 30-33, wherein the glycans are modified.

15 54. The method of claim 53, wherein the glycans are modified by permethylation or conjugation to a peptide.

55. The method of any of claims 1, 11, 26 and 30-33, wherein the method further comprises generating a list of all possible glycoforms.

20

56. The method of any of claims 1, 11, 26 and 30-33, wherein the method further comprises generating a list of possible glycans.

25 57. The method of any of claims 1, 26 and 30-33, wherein the method further comprises removing abundant or nonglycosylated proteins from the sample.

58. The method of claim 57, wherein the abundant or nonglycosylated proteins are albumins or immunoglobulins.

30 59. The method of any of claims 1, 26 and 30-33, wherein the detection limit of the method is less than 5 femtomoles.

60. The method of any of claims 1, 26 and 30-33, wherein the detection limit of the method is less than 1 femtomole.

61. The method of any of claims 1, 26 and 30-33, wherein the method detects low abundance species.

5 62. The method of claim 61, wherein the low abundance species are glycans that contain fucoses, sialic acids, galactoses, mannoses or sulfate groups.

63. The method of any of claims 1, 26 and 30-33, wherein the method is performed in a high-throughput manner.

10

64. The method of claim 63, wherein the method or a portion thereof is performed in a 96-well plate or on a protein-binding membrane.

15

65. The method of claim 64, wherein the 96-well plate comprises a protein-binding membrane.

66. The method of claim 64 or 65, wherein the protein-binding membrane is a polyvinylidene difluoride (PVDF) membrane, C-18 membrane or a nitrocellulose membrane.

20

67. The method of any of claims 1, 26 and 30-33, wherein one or more steps of the method or portions thereof are performed with the use of robotics.

68. The method of any of claims 1, 26 and 30-33, wherein neutral and charged glycans are analyzed separately.

25

69. The method of any of claims 1, 26 and 30-33, wherein the glycoconjugate is a peptide-based glycoconjugate.

30

70. The method of any of claims 1, 26 and 30-33, wherein the glycoconjugate is a lipid-based glycoconjugate.

71. A method of analyzing a sample containing one or more glycoconjugates, which comprise one or more glycans conjugated to a non-saccharide component, comprising:

method of:
(a) analyzing the glycoconjugates to determine the glycosylation sites and glycosylation site occupancy,

(b) separating the glycans from the sample containing one or more glycoconjugates,

(c) analyzing the glycans to characterize the glycans, and

5 (d) determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample,

wherein determining the glycosylation sites and glycosylation site occupancy comprises cleaving the glycans from the non-saccharide components and labeling the non-saccharide components at their glycosylation sites of a first portion of the sample, cleaving
10 the glycans from the non-saccharide components of a second portion of the sample, analyzing the first and second portions of the sample of glycoconjugates and comparing the results.

72. The method of claim 71, wherein determining the glycosylation sites comprises analyzing the non-saccharide components to characterize the non-saccharide components.

15 73. The method of claim 72, wherein the sequence of the non-saccharide components is determined.

74. The method of claim 71, wherein the non-saccharide components of the first portion are
20 labeled with a labeling agent.

75. The method of claim 71, wherein the non-saccharide components of the second portion are unlabeled.

25 76. The method of claim 71, wherein the non-saccharide components of the second portion are labeled.

77. The method of claim 71, wherein the step of determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample comprises generating constraints from
30 the results of step (a) and/or (c) and solving the constraints.

78. The method of claim 77, wherein further constraints are generated from databases containing information about glycans, non-saccharide components, glycoconjugates or a combination thereof; biosynthetic rules or from information about the sample origin.

79. The method of claim 77, wherein the constraints are one or more mathematical equations and are solved by determining the solution of the one or more mathematical equations.

5 80. The method of claim 71, wherein the method further comprises generating a list of all possible glycoforms.

81. The method of claim 71, wherein step (b) comprises cleaving with PNGase F, Endo H, Endo F, hydrazinolysis or alkaline borohydride.

10 82. The method of claim 71, wherein step (b) comprises denaturing the glycoconjugates with a denaturing agent.

83. The method of claim 82, wherein the glycoconjugates are reduced with a reducing agent
15 following their denaturation.

84. The method of claim 83, wherein the glycoconjugates are alkylated with an alkylating agent following their reduction.

20 85. The method of claim 71, wherein the step of analyzing the glycans comprises analyzing the glycans with a mass spectrometric method, an electrophoretic method, NMR, a chromatographic method or some combination thereof.

86. The method of claim 85, wherein the mass spectrometric method is ESI-MS, LC-MS,
25 LC-MS/MS, MALDI-MS, MALDI-MS/MS, MALDI-TOF-MS, MALDI-TOF PSD-MS, MALDI-TOF/TOF-MS, MALDI-TOF/TOF-MS/MS, MALDI-TOF/TOF PSD-MS, MALDI-FTMS, LC-MALDI-TOF/TOF-MS, Nano-LC MALDI-TOF/TOF-MS, Nano-LC MALDI-TOF/TOF PSD-MS, Nano-LC MALDI-TOF/TOF-MS/MS or TANDEM-MS.

30 87. The method of claim 85, wherein the mass spectrometric method is performed in the presence of a thymine derivative and an ion exchange resin.

88. The method of claim 71, wherein the first and second portions of the sample are analyzed with a mass spectrometric method.

89. The method of claim 71, wherein the glycans are analyzed with NMR and MALDI-MS.

90. The method of claim 1 or 89, wherein neutral and charged glycans are analyzed
5 separately.

91. A method of analyzing a sample containing one or more carbohydrates, comprising:

(a) analyzing the carbohydrates with MALDI-MS to determine the monomer
composition and relative abundance of the carbohydrates,

10 (b) analyzing the carbohydrates with NMR to determine the monomer composition
and linkage abundance of the carbohydrates,

(c) and generating constraints from the results of (a) and (b) and solving the
constraints with a computational method.

15 92. The method of claim 91, wherein the carbohydrates are polysaccharides.

93. The method of claim 92, wherein the polysaccharides are glycans.

94. The method of claim 93, wherein the glycans are branched.

20

95. The method of claim 93, wherein the glycans are conjugated to a non-saccharide
component and form one or more glycoconjugates.

96. The method of claim 95, wherein the method further comprises analyzing the non-
25 saccharide components or one or more glycoconjugates or a combination thereof.

97. The method of claim 93 or 96, wherein further constraints are generated from databases
containing information about glycans, non-saccharide components, glycoconjugates or a
combination thereof.

30

98. The method of claim 93 or 96, wherein further constraints are generated from
biosynthetic rules.

99. The method of claim 91, wherein further constraints are generated by analyzing the carbohydrates with another experimental method.

100. The method of claim 99, wherein the other experimental method is a mass
5 spectrometric method that is not MALDI-MS, an electrophoretic method, a different NMR method, a chromatographic method or some combination thereof.

101. The method of claim 91, wherein further constraints are generated from information about the sample origin.

102. The method of claim 91, wherein the constraints are one or more mathematical
10 equations and are solved by determining the solution of the one or more mathematical equations.

103. The method of claim 91, wherein the constraints are solved with a computer program.

104. The method of claim 93, wherein the method further comprises generating a list of all possible glycans.

105. The method of claim 93, wherein neutral and charged glycans are analyzed separately.

106. The method of claim 91, wherein the MALDI-MS is performed in the presence of a thymine derivative and an ion exchange resin.

107. The method of claim 93, wherein a mixture of glycans is analyzed.

108. A method of analyzing a sample containing glycans, comprising:
separating neutral from charged glycans, and
analyzing the neutral and charged glycans separately to characterize the glycans,
30 wherein the glycans are part of glycoconjugates, which comprise one or more glycans conjugated to a non-saccharide component.

109. The method of claim 108, wherein the method further comprises:
denaturing the glycoconjugates,

separating the glycans from the non-saccharide components, and
analyzing the glycans.

110. The method of claim 108, wherein the glycans are analyzed with NMR or MALDI-MS.

111. The method of claim 108, wherein the glycans are analyzed with NMR and MALDI-MS.

112. The method of claim 110 or 111, wherein the MALDI-MS is performed in the presence of a thymine derivative and an ion exchange resin

113. The method of claim 1, 110 or 111, wherein the analyzing further comprises analyzing the glycans with another experimental method.

114. The method of claim 113, wherein the other experimental method is a mass spectrometric method that is not MALDI-MS, an electrophoretic method, a different NMR method, a chromatographic method or some combination thereof.

115. The method of claim 1, 110 or 111, further comprising generating or using a list of all possible glycans.

116. The method of claim 115, wherein the list is based on the results of the analysis of the glycans or from biosynthetic rules or a combination thereof.

117. The method of claim 1, 108, 110 or 111, wherein the glycoconjugates are in solution or immobilized on a solid support.

118. The method of claim 117, wherein the solid support is in a 96-well plate format or comprises a membrane.

119. The method of claim 118, wherein the membrane is a protein-binding membrane.

120. The method of claim 119, wherein the protein-binding membrane is a polyvinylidene difluoride (PVDF) membrane, C-18 membrane or a nitrocellulose membrane.

The method of claim 121, wherein:

121. A method of analyzing a sample containing one or more carbohydrates, comprising:
analyzing the carbohydrates in the presence of a thymine derivative and an ion
exchange resin.

5

122. The method of claim 121, wherein the carbohydrates are polysaccharides.

123. The method of claim 122, wherein the polysaccharides are glycans.

10 124. The method of claim 123, wherein the glycans comprise hyaluronic acid.

125. The method of claim 123, wherein the glycans are part of glycoconjugates, which
comprise one or more glycans conjugated to a non-saccharide component, and wherein the
method further comprises:

15 denaturing the glycoconjugates,
separating the glycans from the non-saccharide components, and
analyzing the glycans.

20 126. The method of claim 121, wherein the thymine derivative is 6-aza-2-thiothymine
(ATT).

127. The method of claim 121, wherein the ion exchange resin is a perfluorinated ion
exchange resin.

25 128. The method of claim 127, wherein the perfluorinated ion exchange resin is NafionTM.

129. The method of claim 121, wherein the carbohydrates are analyzed with NMR or
MALDI-MS.

30 130. The method of claim 121, wherein the carbohydrates are analyzed with NMR and
MALDI-MS.

131. The method of claim 129 or 130, wherein the results from the NMR provide monomer
composition and linkage information for the carbohydrates.

with the glycan structure

132. The method of claim 131, wherein the results are used to generate constraints.

133. The method of claim 132, wherein the constraints are one or more mathematical
5 equations and are solved by determining the solution of the one or more mathematical
equations.

134. The method of claim 129 or 130, wherein the results from the MALDI-MS provide
monomer composition and relative abundance information for the carbohydrates.

10 135. The method of claim 134, wherein the results are used to generate constraints.

136. The method of claim 135, wherein the constraints are one or more mathematical
equations and are solved by determining the solution of the one or more mathematical
15 equations.

137. The method of claim 121, 129 or 130, wherein neutral and charged carbohydrates are
analyzed separately.

20 138. The method of claim 121, 129 or 130, wherein the analyzing further comprises
analyzing the carbohydrates with another experimental method.

139. The method of claim 138, wherein the other experimental method is a mass
spectrometric method, an electrophoretic method, NMR, a chromatographic method or some
25 combination thereof.

140. The method of claim 123, 129 or 130, further comprising generating or using a list of
all possible glycans.

30 141. The method of claim 140, wherein the list is based on the results of the analysis of the
glycans or from biosynthetic rules or a combination thereof.

142. The method of claim 121, 129 or 130, wherein the carbohydrates are in solution or
immobilized on a solid support.

143. The method of claim 142, wherein the solid support is in a 96-well plate format or comprises a membrane.

5 144. The method of claim 143, wherein the membrane is a protein-binding membrane.

145. The method of claim 144, wherein the protein-binding membrane is a polyvinylidene difluoride (PVDF) membrane, C-18 membrane or a nitrocellulose membrane.

10 146. The method of claim 121, 129 or 130, wherein the carbohydrates are in a mixture of carbohydrates.

147. A method of analyzing a sample containing glycans, comprising:

analyzing a first portion of the sample, wherein the glycans have been removed,
15 analyzing a second portion of the sample, wherein the second portion of the sample contains intact glycoconjugates, which comprise one or more glycans conjugated to a non-saccharide component,
and analyzing a third portion of the sample, wherein the third portion of the sample contains glycans.

20

148. The method of claim 147, wherein the second portion of the sample containing intact glycoconjugates is analyzed with a method, comprising:

(a) analyzing the glycoconjugates to characterize the glycoconjugates,
(b) analyzing the non-saccharide components of the glycoconjugates to characterize
25 the non-saccharide components,
(c) separating the glycans from the sample containing one or more glycoconjugates,
(d) analyzing the glycans to characterize the glycans, and
(e) determining the identity and quantity of all of the glycoforms of the
glycoconjugates in the sample with the results obtained from steps (a), (b) and (d) and a
30 computational method.

149. The method of claim 147, wherein the second portion of the sample containing intact glycoconjugates is analyzed with a method, comprising:

(a) analyzing the glycoconjugates to determine the glycosylation sites and glycosylation site occupancy,

(b) separating the glycans from the sample containing one or more glycoconjugates,

(c) analyzing the glycans to characterize the glycans, and

(d) determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample,

wherein determining the glycosylation sites and glycosylation site occupancy comprises cleaving the glycans from the non-saccharide components and labeling the non-saccharide components at their glycosylation sites of a first portion of the sample, cleaving the glycans from the non-saccharide components of a second portion of the sample, analyzing the first and second portions of the sample of glycoconjugates and comparing the results.

150. The method of claim 147, wherein the second portion of the sample containing intact glycoconjugates is analyzed with a method, comprising:

(a) denaturing the glycoconjugates,

(b) separating the glycans from the non-saccharide components of the glycoconjugates,

(c) analyzing the glycans with MALDI-MS to determine the monosaccharide composition and relative abundance of the glycans,

(d) analyzing the glycans with NMR to determine the monosaccharide composition and linkage abundance of the glycans, and

(e) generating constraints from the results of (c) and (d) and solving the constraints with a computational method.

151. The method of any of claims 147-150, wherein the second portion of the sample containing intact glycoconjugates is analyzed with a method, comprising:

denaturing the glycoconjugates,

separating the glycans from the non-saccharide components of the glycoconjugates,

separating neutral from charged glycans, and

analyzing the neutral and charged glycans separately to characterize the glycans.

152. The method of any of claims 147-150, wherein the second portion of the sample containing intact glycoconjugates is analyzed with a method, comprising:

denaturing the glycoconjugates,

separating the glycans from the non-saccharide components of the glycoconjugates,
and

analyzing the glycans in the presence of a thymine derivative and an ion exchange
resin.

5

153. The method of any of claims 147-150, wherein the third portion of the sample
containing glycans is analyzed with a method, comprising:

(a) analyzing the glycans with MALDI-MS to determine the monosaccharide
composition and relative abundance of the glycans,

10

(b) analyzing the glycans with NMR to determine the monosaccharide composition
and linkage abundance of the glycans, and

(c) generating constraints from the results of (a) and (b) and solving the constraints
with a computational method.

15

154. The method of any of claims 147-150, wherein the third portion of the sample
containing glycans is analyzed with a method, comprising:

separating neutral from charged glycans, and

analyzing the neutral and charged glycans separately to characterize the glycans.

20

155. The method of any of claims 147-150, wherein the third portion of the sample
containing glycans is analyzed with a method, comprising:

analyzing the glycans in the presence of a thymine derivative and an ion exchange
resin.

25

156. The method of any of claims 147-155, wherein the glycans of the third portion of the
sample are not part of intact glycoconjugates.

157. The method of any of claims 147-155, wherein the glycans of the third portion of the
sample are part of intact glycoconjugates, which comprise one or more glycans conjugated to
a non-saccharide component, and the method further comprises:

30

denaturing the glycoconjugates, and

separating the glycans from the non-saccharide components of the glycoconjugates.

158. The method of any of claims 1-157, wherein the sample is a sample comprising one or more glycoconjugates, one or more cells, a tissue or body fluid from a subject, or the sample is a batch of glycoconjugates.
- 5 159. The method of claim 1-157, wherein the sample is a sample comprising one or more glycoconjugates.
160. The method of any of claims 1-157, wherein the method is a method of analyzing the purity of the sample.
- 10 161. The method of any of claims 1-157, wherein more than one sample containing glycans are analyzed.
162. The method of claim 161, wherein the more than one samples are two or more batches
- 15 of glycoconjugates.
163. The method of claim 161, wherein the more than one samples are contained in a 96-well plate or on a protein-binding membrane.
- 20 164. A method of generating a glycoconjugate library, wherein the glycoconjugate comprises one or more glycans conjugated to a non-saccharide component, comprising:
- cleaving the non-saccharide component of the glycoconjugate in a sample and labeling the non-saccharide component fragments generated with a labeling agent in order to generate a glycoconjugate library, and
- 25 cleaving the glycans from the non-saccharide components and labeling the non-saccharide components in the sample at the glycosylation sites.
165. The method of claim 164, wherein the labeling agent is an isotope of C, N, H, S or O.
- 30 166. The method of claim 165, wherein the labeling agent is ^{18}O .
167. The method of claim 165, wherein the labeling agent is ^2H .

168. The method of claim 164, wherein the method further comprises analyzing the fragments generated from the cleavage of the glycoconjugates.

169. The method of claim 168, wherein the fragments are analyzed with ESI-MS, LC-MS,
5 LC-MS/MS, MALDI-MS, MALDI-MS/MS, MALDI-TOF-MS, MALDI-TOF PSD-MS, MALDI-TOF/TOF-MS, MALDI-TOF/TOF-MS/MS, MALDI-TOF/TOF PSD-MS, MALDI-FTMS, LC-MALDI-TOF/TOF-MS, Nano-LC MALDI-TOF/TOF-MS, Nano-LC MALDI-TOF/TOF PSD-MS, Nano-LC MALDI-TOF/TOF-MS/MS or TANDEM-MS.

10 170. The method of claim 164, wherein the glycoconjugate is a peptide-based glycoconjugate.

171. The method of claim 170, wherein the analyzing results in the characterization of the glycosylation sites, the peptides and the glycans of the peptide-based glycoconjugate.

15 172. A library generated with the method of any of claims 164-167.

173. A method of analyzing a sample containing glycoconjugates, comprising:
modifying the glycoconjugates in the sample, and
20 comparing the modified glycoconjugates with the library of claim 164.

174. The method of claim 173, wherein the step of modifying the glycoconjugates comprises cleaving the glycoconjugates to generate glycoconjugate fragments.

25 175. The method of claim 174, wherein the glycoconjugates are cleaved by cleaving the non-saccharide component of the glycoconjugates, cleaving the glycans of the glycoconjugates, cleaving the glycans from the non-saccharide components of the glycoconjugates or a combination thereof to generate fragments.

30 176. The method of claim 174, wherein the step of comparing includes mixing the glycoconjugate fragments with the library and determining the ratios of the glycoconjugate fragments to the library.

177. The method of claim 176, wherein known proportions of samples of the glycoconjugates and fragments are mixed with the library.

178. The method of claim 173, wherein the sample containing glycoconjugates is a batch of glycoconjugates.

179. The method of claim 173, wherein the sample containing glycoconjugates is a sample of one or more cells, a tissue or body fluid from a subject.

180. A method of generating a list of properties, comprising:

determining one or more properties of a sample with the method of any of claims 1-157, and

recording a value for the one or more properties to generate a list, wherein the value of the one or more properties is recorded in a computer-generated data structure.

181. The method of claim 180, wherein the one or more properties comprises the number of one or more types of monosaccharides of a glycan in the sample.

182. The method of claim 180, wherein the one or more properties comprises the mass of a glycan in the sample.

183. The method of claim 180, wherein the one or more properties comprises the quantity of a glycan in the sample.

184. The method of any of claims 181-183, wherein the glycan is a permethylated glycan.

185. The method of any of claims 181-183, wherein the glycan is an unmodified glycan.

186. The method of any of claims 181-183, wherein the glycan is part of a glycoconjugate.

187. The method of claim 180, wherein the sample contains a glycoconjugate.

188. The method of claim 187, wherein the glycoconjugate is a peptide-based glycoconjugate, and wherein the one or more properties comprises the mass of the peptide of the peptide-based glycoconjugate.

5 189. The method of claim 187, wherein the glycoconjugate is a lipid-based glycoconjugate, and wherein the one or more properties comprises the mass of the lipid of the lipid-based glycoconjugate.

190. The method of claim 187, wherein the one or more properties comprises the mass of the
10 glycoconjugate.

191. A database, tangibly embodied in a computer-readable medium, for storing information descriptive of one or more carbohydrates, the database comprising:

one or more data units corresponding to the one or more carbohydrates, each of the
15 data units including an identifier that includes one or more fields, each field for storing a value corresponding to one or more properties of the carbohydrates, wherein the value corresponding to one or more properties of the carbohydrates is determined with a method, comprising:

(a) analyzing the carbohydrates with MALDI-MS to determine the monomer
20 composition and relative abundance of the carbohydrates, and

(b) analyzing the carbohydrates with NMR to determine the monomer composition and linkage abundance of the carbohydrates.

192. The database of claim 191, wherein the method further comprises:

25 (c) generating constraints from the results of (a) and (b) and solving the constraints with a computational method.

193. The database of claim 191, wherein the carbohydrates are polysaccharides.

30 194. The database of claim 191, wherein the polysaccharides are glycans.

195. A database, tangibly embodied in a computer-readable medium, for storing information descriptive of one or more carbohydrates, the database comprising:

one or more data units corresponding to the one or more carbohydrates, each of the data units including an identifier that includes one or more fields, each field for storing a value corresponding to one or more properties of the carbohydrates, wherein the value corresponding to one or more properties of the carbohydrates is determined with a method, comprising:

separating neutral from charged carbohydrates, and
analyzing the neutral and charged carbohydrates separately to characterize the glycans.

196. The database of claim 195, wherein the carbohydrates are polysaccharides.

197. The database of claim 195, wherein the polysaccharides are glycans.

198. A database, tangibly embodied in a computer-readable medium, for storing information descriptive of one or more carbohydrates, the database comprising:

one or more data units corresponding to the one or more carbohydrates, each of the data units including an identifier that includes one or more fields, each field for storing a value corresponding to one or more properties of the carbohydrates, wherein the value corresponding to one or more properties of the carbohydrates is determined with a method, comprising:

analyzing the carbohydrates in the presence of a thymine derivative and an ion exchange resin.

199. The database of claim 198, wherein the carbohydrates are polysaccharides.

200. The database of claim 198, wherein the polysaccharides are glycans.

201. The database of any of claims 194, 197 and 200, wherein the glycans are part of intact glycoconjugates, which comprise one or more glycans conjugated to a non-saccharide component, and the method further comprises:

denaturing the glycoconjugates, and
separating the glycans from the non-saccharide components of the glycoconjugates.

202. A database, tangibly embodied in a computer-readable medium, for storing information descriptive of one or more glycoconjugates, the database comprising:

one or more data units corresponding to the one or more glycoconjugates, each of the data units including an identifier that includes one or more fields, each field for storing a value corresponding to one or more properties of the glycoconjugates, wherein the value corresponding to one or more properties of the glycoconjugates is determined with a method, comprising:

- (a) analyzing the glycoconjugates to characterize the glycoconjugates,
- (b) analyzing the non-saccharide components of the glycoconjugates to characterize the non-saccharide components,
- (c) separating the glycans from the sample containing one or more glycoconjugates, and
- (d) analyzing the glycans to characterize the glycans.

203. The database of claim 202, wherein the method further comprises:

- (e) determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample with the results obtained from steps (a), (b) and (d) and a computational method.

204. A database, tangibly embodied in a computer-readable medium, for storing information descriptive of one or more glycoconjugates, the database comprising:

one or more data units corresponding to the one or more glycoconjugates, each of the data units including an identifier that includes one or more fields, each field for storing a value corresponding to one or more properties of the glycoconjugates, wherein the value corresponding to one or more properties of the glycoconjugates is determined with a method, comprising:

- (a) analyzing the glycoconjugates to determine the glycosylation sites and glycosylation site occupancy,
- (b) separating the glycans from the sample containing one or more glycoconjugates, and
- (c) analyzing the glycans to characterize the glycans,

wherein determining the glycosylation sites and glycosylation site occupancy comprises cleaving the glycans from the non-saccharide components and labeling the non-saccharide components at their glycosylation sites of a first portion of the sample, cleaving

the glycans from the non-saccharide components of a second portion of the sample, analyzing the first and second portions of the sample of glycoconjugates and comparing the results.

205. The database of claim 204, wherein the method further comprises:

5 (d) determining the identity and quantity of all of the glycoforms of the glycoconjugates in the sample.

206. A database, tangibly embodied in a computer-readable medium, for storing information descriptive of one or more glycoconjugates, the database comprising:

10 one or more data units corresponding to the one or more glycoconjugates, each of the data units including an identifier that includes one or more fields, each field for storing a value corresponding to one or more properties of the glycoconjugates, wherein the value corresponding to one or more properties of the glycoconjugates is determined with a method, comprising:

15 (a) denaturing the glycoconjugates,

(b) separating the glycans from the non-saccharide components of the glycoconjugates,

(c) analyzing the glycans with MALDI-MS to determine the monosaccharide composition and relative abundance of the glycans, and

20 (d) analyzing the glycans with NMR to determine the monosaccharide composition and linkage abundance of the glycans.

207. The database of claim 206, wherein the method further comprises:

25 (e) generating constraints from the results of (c) and (d) and solving the constraints with a computational method.

208. A database, tangibly embodied in a computer-readable medium, for storing information descriptive of one or more glycoconjugates, the database comprising:

30 one or more data units corresponding to the one or more glycoconjugates, each of the data units including an identifier that includes one or more fields, each field for storing a value corresponding to one or more properties of the glycoconjugates, wherein the value corresponding to one or more properties of the glycoconjugates is determined with a method, comprising:

denaturing the glycoconjugates,

separating the glycans from the non-saccharide components of the glycoconjugates, separating neutral from charged glycans, and analyzing the neutral and charged glycans separately to characterize the glycans.

- 5 209. A database, tangibly embodied in a computer-readable medium, for storing information descriptive of one or more glycoconjugates, the database comprising:

one or more data units corresponding to the one or more glycoconjugates, each of the data units including an identifier that includes one or more fields, each field for storing a value corresponding to one or more properties of the glycoconjugates,

- 10 wherein the value corresponding to one or more properties of the glycoconjugates is determined with a method, comprising:

denaturing the glycoconjugates,

separating the glycans from the non-saccharide components of the glycoconjugates,

and

- 15 analyzing the glycans in the presence of a thymine derivative and an ion exchange resin.

210. A method of analyzing the total glycome of a sample, comprising:

- (a) analyzing all of the glycans of the sample, and
20 (b) determining a profile of the glycans of the sample.

211. The method of claim 210, wherein the composition of the glycans in the sample is determined.

- 25 212. The method of claim 210, wherein the structures of the glycans in the sample are determined.

213. The method of claim 210, wherein step (a) includes quantifying the glycans using calibration curves based on known glycan standards.

30

214. The method of claim 210, wherein the profile of the glycans is a spectrum of monosaccharide composition and relative abundance of the glycans.

215. The method of claim 210, wherein the glycans of the sample are analyzed with a mass spectrometric method, an electrophoretic method, NMR, a chromatographic method or a combination thereof.

216. The method of claim 215, wherein the analysis is performed with ESI-MS, LC-MS, LC-MS/MS, MALDI-TOF-MS, MALDI-MS/MS, MALDI-FTMS, TANDEM-MS, NMR, HPLC or CE.

217. The method of claim 216, wherein the analysis is performed with MALDI-MS or MALDI-FTMS.

218. The method of claim 210, wherein the analysis is performed with electrophoresis, microfluidic devices or nanofluidic devices.

219. The method of claim 215, wherein the glycans are further analyzed with another experimental method.

220. The method of claim 219, wherein the other experimental method provides linkage information.

221. The method of claim 220, wherein the other experimental method is LC-MS, LC-MS/MS, CE-LIF or NMR.

222. The method of claim 219, wherein the other experimental method comprises the use of one or more glycan-degrading enzymes.

223. The method of claim 210, wherein the glycans of the sample are analyzed with a method, comprising:

analyzing the glycans with MALDI-MS to determine the monosaccharide composition and relative abundance of the glycans, and
analyzing the glycans with NMR to determine the monosaccharide composition and linkage abundance of the glycans.

224. The method of claim 223, wherein the method further comprises:

sample, generating constraints from the results of the analysis with MALDI-MS and NMR and
solving the constraints with a computational method.

225. The method of claim 210, 223 or 224 wherein the glycans of the sample are analyzed
5 with a method, comprising:

separating neutral from charged glycans, and
analyzing the neutral and charged glycans separately to characterize the glycans.

226. The method of claim 210, 223 or 224 wherein the glycans of the sample are analyzed
10 with a method, comprising:

analyzing the glycans in the presence of a thymine derivative and an ion exchange
resin.

227. The method of claim 226, wherein the thymine derivative is 6-aza-2-thiothymine (ATT)
15 and the ion exchange resin is a perfluorinated ion exchange resin.

228. The method of claim 227, wherein the perfluorinated ion exchange resin is NafionTM.

229. The method of any of claims 210 and 223-226, wherein the glycans are part of intact
20 glycoconjugates, which comprise one or more glycans conjugated to non-saccharide
components.

230. The method of claim 229, wherein the method further comprises:
separating the glycans from the non-saccharide components of the glycoconjugates.
25

231. The method of claim 230, wherein the glycans are separated by cleavage with an
enzymatic method or a chemical method.

232. The method of claim 231, wherein the enzymatic method includes the use of PNGase F,
30 Endo H or Endo F.

233. The method of claim 231, wherein the chemical method includes hydrazinolysis or
alkaline borohydride.

234. The method of claim 231, wherein the cleavage is performed in a 96-well plate, on a protein-binding membrane or in solution.

235. The method of claim 231, wherein the cleavage is performed with the use of robotics or manually.

236. The method of claim 230, wherein the method further comprises purification of the glycans.

237. The method of claim 236, wherein the purification is performed in a 96-well plate.

238. The method of claim 236, wherein the purification is performed using individual purification columns or cartridges.

239. The method of claim 236, wherein the purification is performed with solid phase extraction cartridges.

240. The method of claim 236, wherein the purification is performed with the use of robotics or manually.

241. The method of claim 210, 223 or 224 wherein the entire sample is analyzed.

242. The method of claim 210, 223 or 224 wherein a fraction of the sample is analyzed.

243. The method of claim 210, 223 or 224 wherein the method further comprises fractionation of a part of the sample.

244. The method of claim 243, wherein the fractionation is based on charge, size, molecular weight, binding properties, acidity, basicity, pI, hydrophobicity or hydrophilicity.

245. The method of claim 243, wherein the fractionation is performed using solid supports with immobilized proteins, organic molecules, inorganic molecules, lipids, carbohydrates or nucleic acids; filters or resins.

246. The method of claim 243, wherein glycans of the fractionated part of the sample are the glycans that are analyzed.

247. The method of claim 243, wherein the fractionated part of the sample is isolated.

248. The method of claim 243, wherein a fraction of the sample is removed and it is the remaining fraction that is analyzed.

249. The method of claim 248, wherein the fraction of the sample removed contains acidic glycans.

250. The method of claim 248, wherein the fraction of the sample removed contains neutral glycans.

251. The method of claim 248, wherein the fraction of the sample removed contains high abundance proteins.

252. The method of claim 251, wherein the high abundance proteins are albumins or immunoglobulins.

253. The method of claim 252, wherein the high abundance proteins are immunoglobulins.

254. The method of claim 248, wherein the fraction of the sample removed does not contain high abundance proteins.

255. The method of claim 230, wherein the method further comprises:
denaturing the glycoconjugates.

256. The method of claim 255, wherein the method further comprises reducing and/or alkylating the glycoconjugates.

257. The method of claim 229, wherein the glycans that are part of intact glycoconjugates are analyzed with a method, comprising:

analyzing the glycoconjugates to characterize the glycoconjugates,

analyzing the non-saccharide components of the glycoconjugates to characterize the
non-saccharide components,
separating the glycans from the sample, and
analyzing the glycans to characterize the glycans.

5

258. The method of claim 257, wherein the method further comprises:

determining the identity and quantity of all of the glycoforms of the glycoconjugates
in the sample with the results obtained from the analysis and a computational method.

10 259. The method of claim 229, wherein the glycans that are part of intact glycoconjugates
are analyzed with a method, comprising:

analyzing the glycoconjugates to determine the glycosylation sites and glycosylation
site occupancy,

separating the glycans from the sample, and

15 analyzing the glycans to characterize the glycans,

wherein determining the glycosylation sites and glycosylation site occupancy
comprises cleaving the glycans from the non-saccharide components and labeling the non-
saccharide components at their glycosylation sites of a first portion of the sample, cleaving
the glycans from the non-saccharide components of a second portion of the sample, analyzing
20 the first and second portions of the sample and comparing the results.

260. The method of claim 259, wherein the method further comprises:

determining the identity and quantity of all of the glycoforms of the glycoconjugates
in the sample.

25

261. The method of claim 210, 223 or 224 wherein the method further comprises identifying
a pattern by performing a pattern analysis on the results from (a) using a computational
method.

30 262. The method of claim 261, wherein the computational method is an iterative
computational method.

263. The method of claim 262, wherein the iterative computational method determines the
glycoforms in the sample.

Small State, son, b

Small State, son, b

264. The method of claim 210, 223 or 224 wherein the sample is a batch of glycoconjugates.

265. The method of claim 210, 223 or 224 wherein the sample is a sample of one or more
5 cells, a tissue or body fluid.

266. The method of claim 261, wherein the method further includes recording the pattern in
a computer-generated data structure.

10 267. The method of claim 261, wherein the pattern of the glycans provides diagnostic or
prognostic information.

268. The method of claim 261, wherein the method further comprises associating the pattern
with one or more patterns of one or more samples of known origin.

15

269. The method of claim 261, wherein the pattern of the glycans provides information about
the sample origin.

270. The method of claim 261, wherein the sample is from a subject and the pattern of the
20 glycans provides information about the subject's state.

271. The method of claim 270, wherein the subject's state is a diseased state.

272. The method of claim 261, wherein the identified pattern is compared to the pattern of at
25 least one other sample.

273. The method of claim 272, wherein the at least one other sample is a batch of
glycoconjugates.

30 274. The method of claim 272, wherein the at least one other sample is a sample from a
healthy or diseased individual.

275. The method of claim 261, wherein the identified pattern is compared to another pattern.

276. The method of claim 275, wherein the other pattern is a known pattern.

277. The method of claim 275, wherein the other pattern is an unknown pattern.

5 278. The method of claim 275, wherein the other pattern is a pattern that represents a batch of glycoconjugates.

279. The method of claim 276, wherein the method is a method to assess the purity of a batch of glycoconjugates and the known pattern represents a batch of glycoconjugates of
10 known purity.

280. The method of claim 275, wherein the other pattern is a pattern that represents a diseased or healthy state.

15 281. The method of claim 280, wherein the diseased state is associated with cancer.

282. The method of claim 281, wherein the cancer is prostate cancer, melanoma, bladder cancer, breast cancer, lymphoma, ovarian cancer, lung cancer, colorectal cancer or head and neck cancer.

20 283. The method of claim 280, wherein the diseased state is associated with an immunological disorder.

284. The method of claim 280, wherein the diseased state is associated with a
25 neurodegenerative disease.

285. The method of claim 284, wherein the neurodegenerative disease is a transmissible spongiform encephalopathy, Alzheimer's disease or neuropathy.

30 286. The method of claim 280, wherein the diseased state is associated with inflammation.

287. The method of claim 280, wherein the diseased state is associated with rheumatoid arthritis.

288. The method of claim 280, wherein the diseased state is associated with cystic fibrosis.

289. The method of claim 280, wherein the diseased state is associated with an infection.

5 290. The method of claim 289, wherein the infection is viral or bacterial.

291. The method of claim 280, wherein the diseased state is associated with a congenital disorder.

10 292. The method of claim 276, wherein the method is a method of monitoring prognosis and the known pattern is associated with the prognosis of a disease.

293. The method of claim 276, wherein the method is a method of monitoring drug treatment and the known pattern is associated with the drug treatment.

15

294. The method of claim 210, wherein the sample is a sample of serum, plasma, blood, urine, saliva, sputum, tears, CSF, seminal fluid, feces, tissues or cells.

20 295. The method of claim 268, wherein the method further comprises validating the association of the pattern with the one or more patterns of the one or more samples of known origin.

296. A method of generating a glycoprofile according to the method of any of claims 210-260.

25

297. A method of creating a database of glycoprofiles, comprising:
generating a glycoprofile of a sample according to the method of claim 296, and
recording one or more values corresponding to the glycoprofile in a computer-generated data structure.

30

298. The database created according to the method of claim 297.

299. A method of determining a glycome pattern, comprising:

obtaining a glycoprofile of total glycans of a sample containing glycans according to the method of any of claims 210-260,

identifying features of the glycoprofile,

generating data sets based on the features of the glycoprofile,

5 identifying a pattern in the data sets,

and determining whether or not the pattern is associated with a known sample or diseased state.

300. The method of claim 299, wherein the sample containing glycans is obtained from a
10 subject.

301. The method of claim 300, wherein the subject has a disease or condition.

302. The method of claim 299, wherein the step of determining the glycoprofile includes
15 obtaining more than one glycoprofile spectra.

303. The method of claim 302, wherein one of the spectra is of acidic glycans.

304. The method of claim 302, wherein one of the spectra is of neutral glycans.
20

305. The method of claim 302, wherein one spectra is of acidic glycans and another spectra is of neutral glycans.

306. The method of claim 299, wherein the feature identified is the presence of one or more
25 glycans, the absence of one or more glycans, the relative amount of one or more glycans, the combination of two or more classes of glycans, the presence of a specific glycan motif, the absence of a specific glycan motif, the relative amount of a specific glycan motif, the presence of one or more monosaccharides in a glycan, the absence of one or more monosaccharides in a glycan, the relative amount of one or more monosaccharides in a
30 glycan or the bond between monosaccharides.

307. The method of claim 299, wherein the pattern is identified by linear discriminant, nearest neighbor, statistical classifier, neural net, decision tree, decision rules or association rules analysis.

The method of claim 299,

The method of claim 299,

308. The method of claim 299, wherein the glycoprofile is generated from determining the glycosylation site occupancy of glycoconjugates in the sample.

5 309. The method of claim 299, wherein the glycoprofile is determined by identifying and quantifying all of the glycans of the sample.

310. The method of claim 309, wherein all of the glycans are identified and quantified by solving constraints with a computational method.

10

311. A method of determining a glycome pattern of a sample, comprising:
determining the glycoprofile of the sample according to the method of any of claims 210-260,

15

extracting one or more features of the glycoprofile,
analyzing the one or more features,
and validating the glycome pattern.

20

312. The method of claim 311, wherein the one or more features is the presence or absence of a specific glycan, an amount of a specific glycan or a combination of specific glycans.

313. The method of claim 311, wherein the one or more features is a ratio between two or more glycans.

25

314. The method of claim 311, wherein the one or more features is the range of amounts of one or more glycans.

315. The method of claim 311, wherein the one or more features is the range of ratios between two or more glycans.

30

316. The method of claim 311, wherein the sample is a batch of glycoconjugates.

317. The method of claim 311, wherein the sample is a sample of body fluid.

318. The method of claim 317, wherein the sample of body fluid is from a subject with a disease or condition.

319. The method of claim 317, wherein the sample of body fluid is from a subject that is
5 undergoing treatment for a disease.

320. The method of claim 317, wherein the sample of body fluid is from a healthy subject.

321. The method of claim 317, wherein the sample of body fluid is from a pregnant woman.
10

322. The method of claim 311, wherein the glycoprofile is generated from determining the glycosylation site occupancy of glycoconjugates in the sample.

323. The method of claim 311, wherein the glycoprofile is determined by identifying and
15 quantifying all of the glycans in the sample.

324. The method of claim 323, wherein the glycans are identified and quantified by solving constraints with a computational method.

20 325. A method of generating a glycome pattern according to the method of claim 299.

326. A method of generating a glycome pattern according to the method of claim 311.

327. A method of creating a database of glycome patterns generated by the method of claim
25 325 or 326.

328. The database created according to claim 327.

1/64

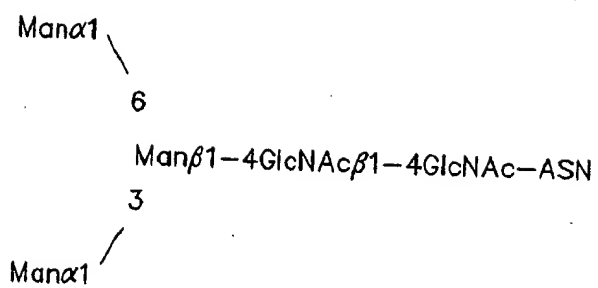


FIG. 1

2/64

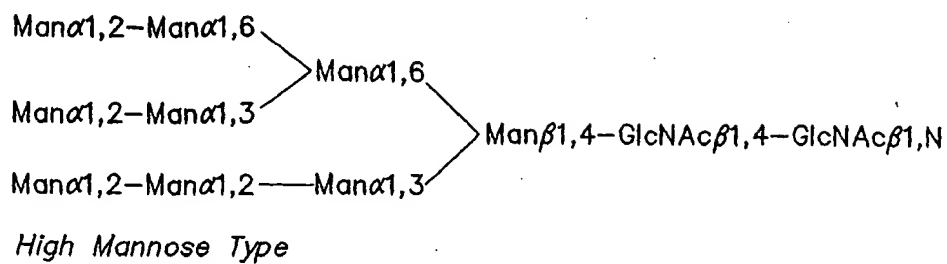


FIG. 2A

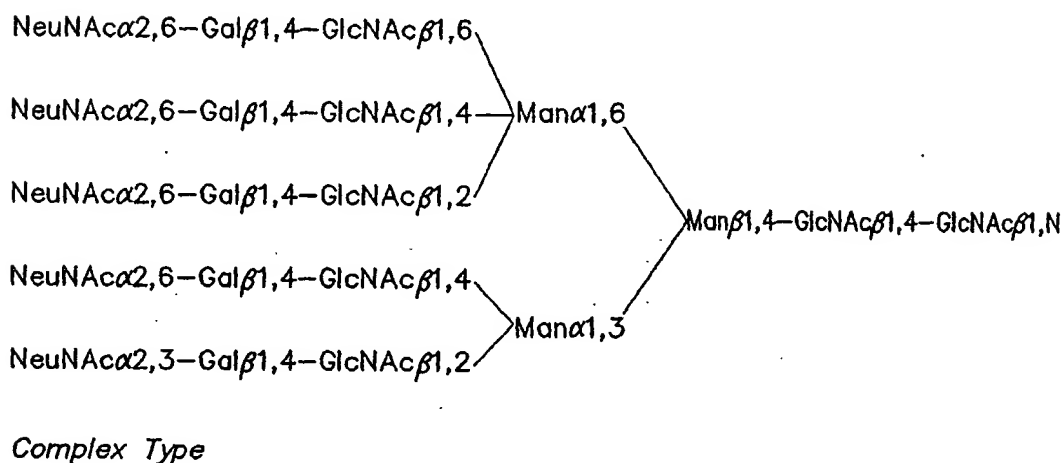


FIG. 2B

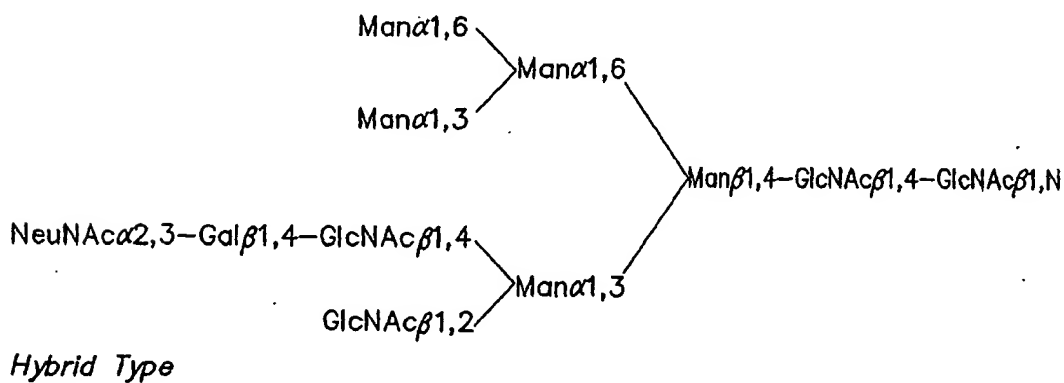


FIG. 2C

3/64

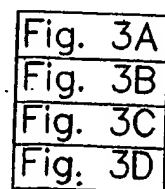
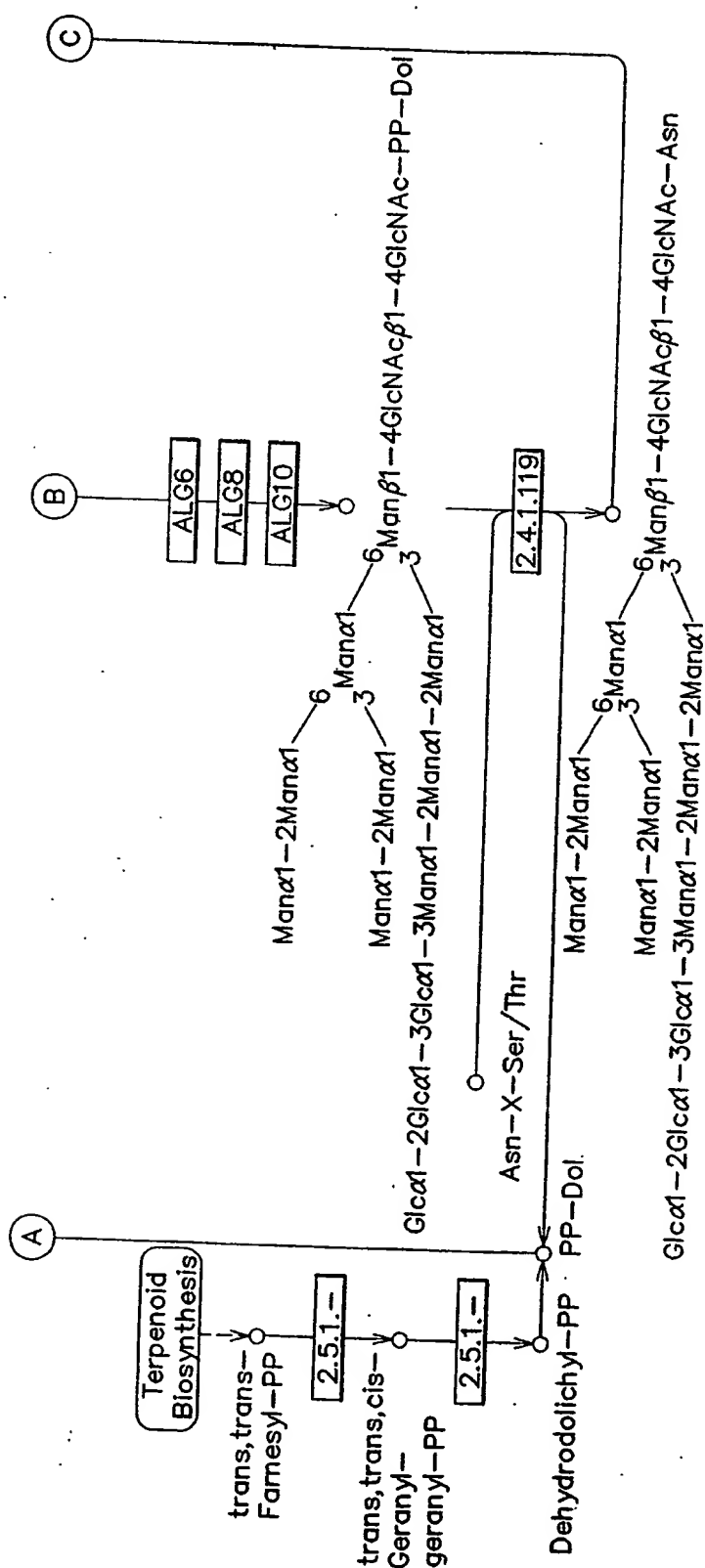


FIG. 3A



5/64

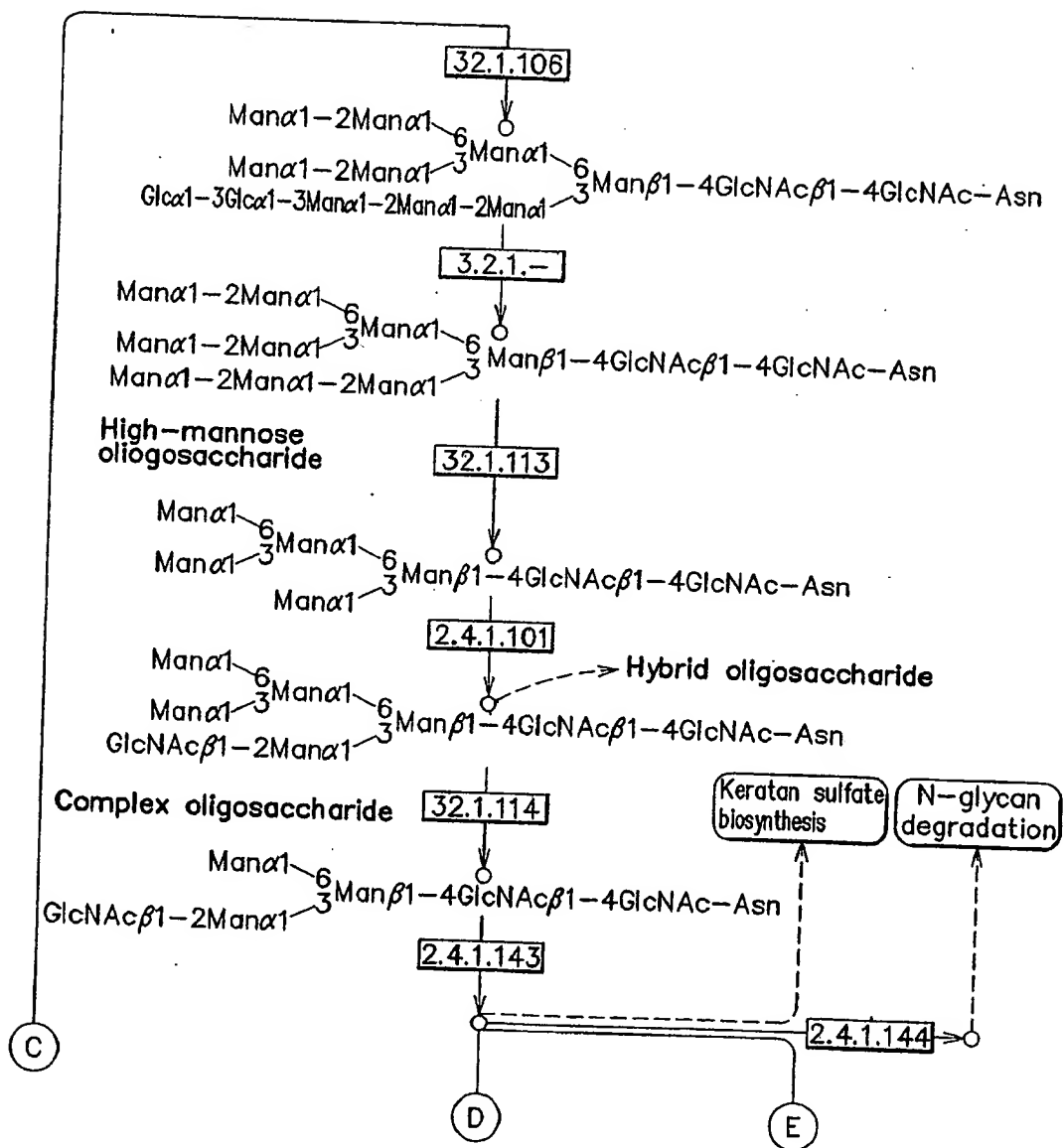


FIG. 3C



FIG. 3D

7/64

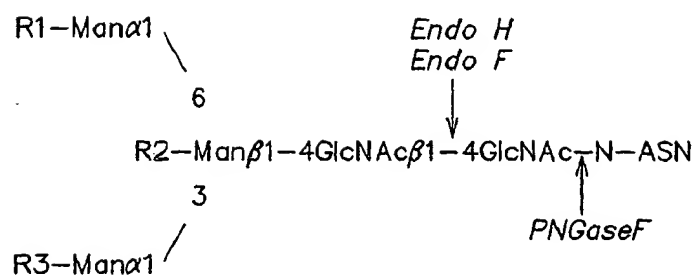


FIG. 4

8/64

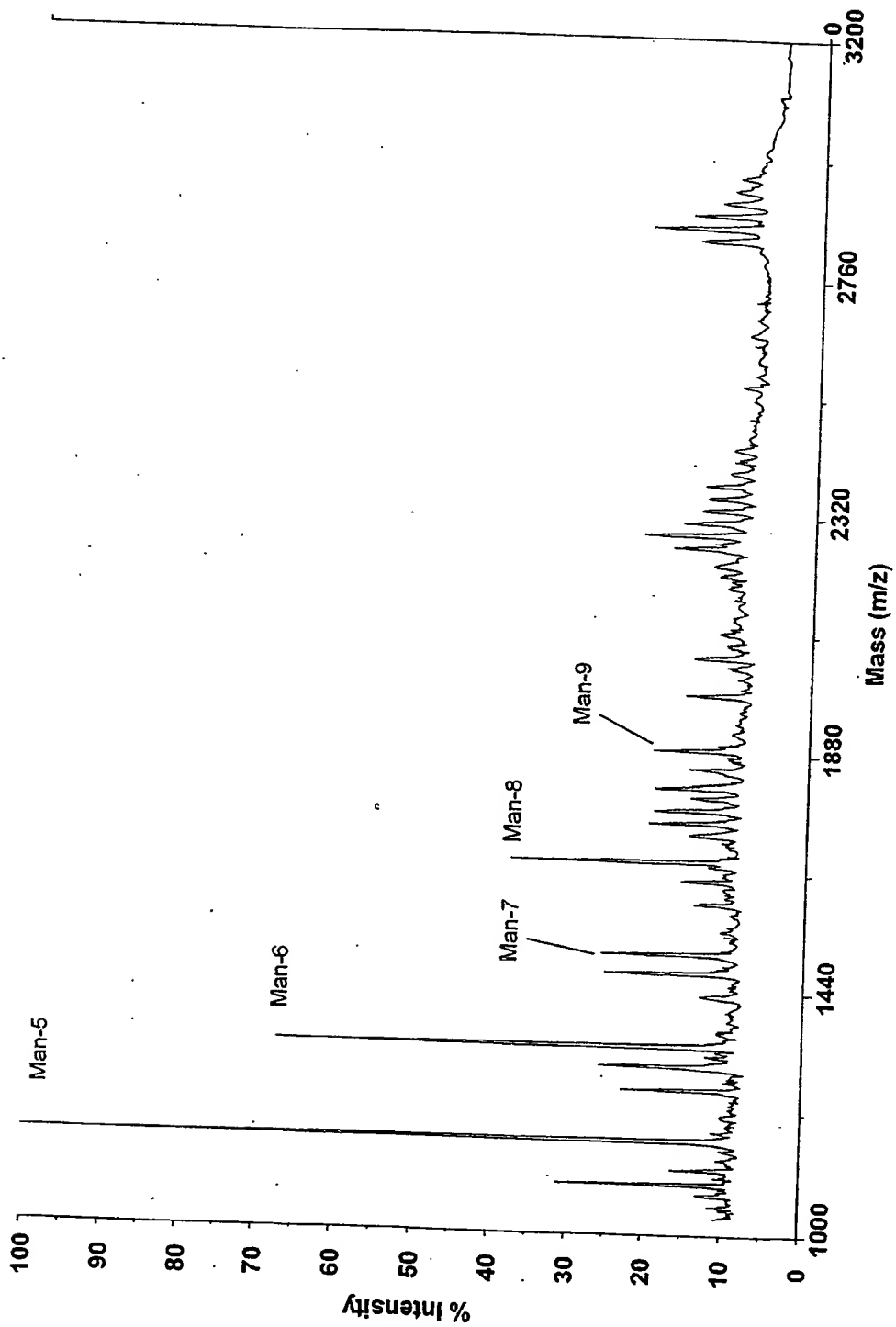


FIG. 5A

9/64

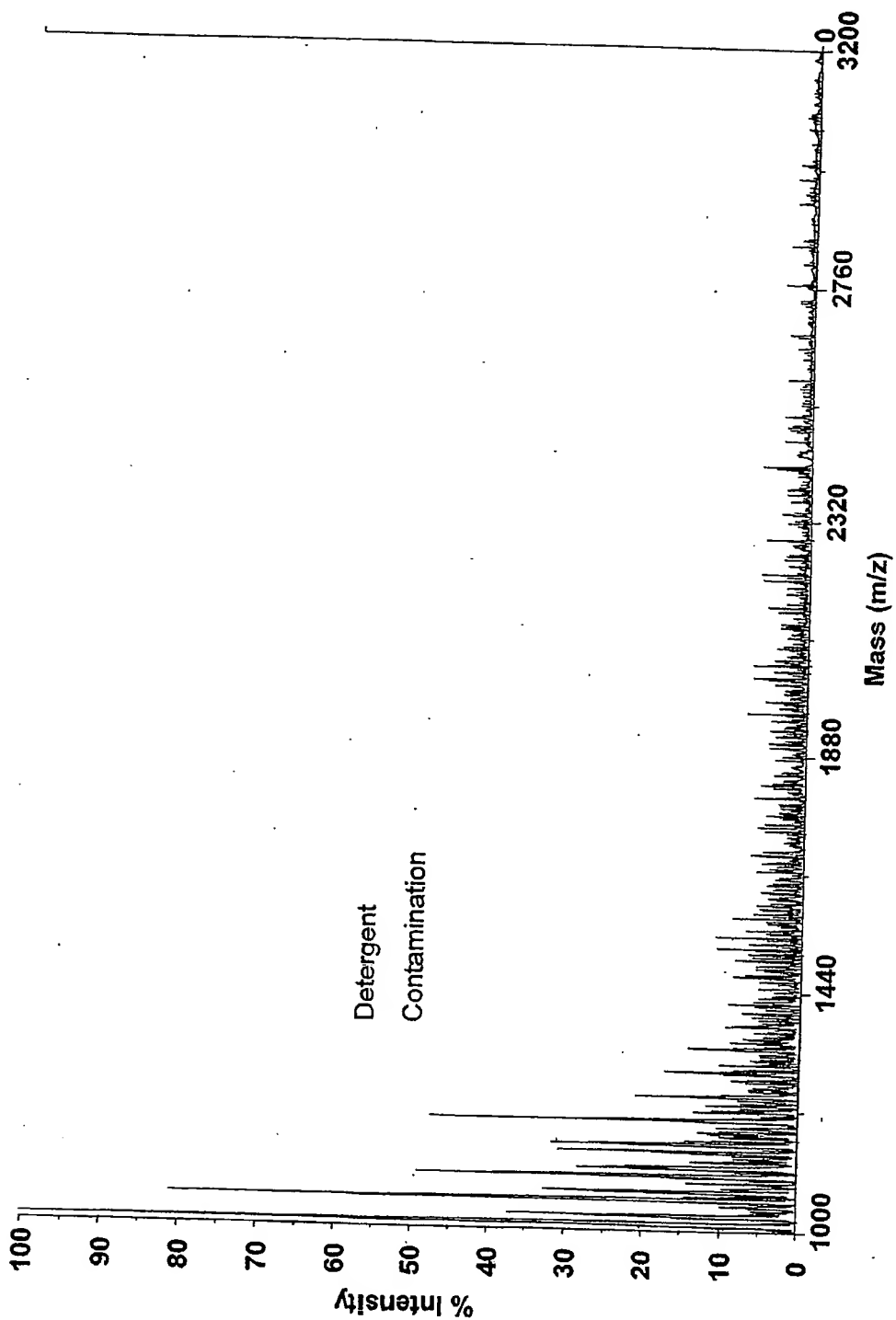


FIG. 5B

10/64

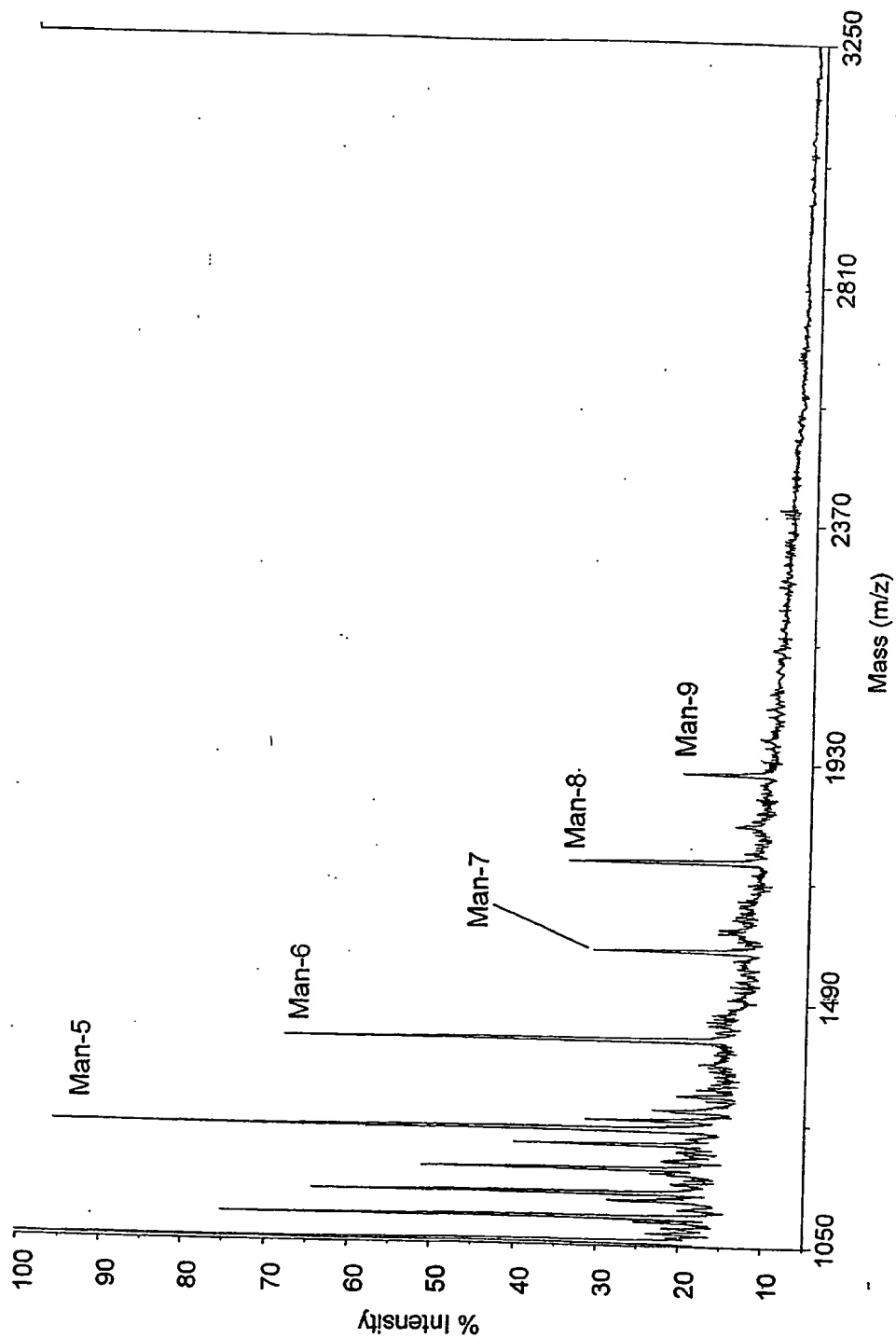


FIG. 5C

11/64

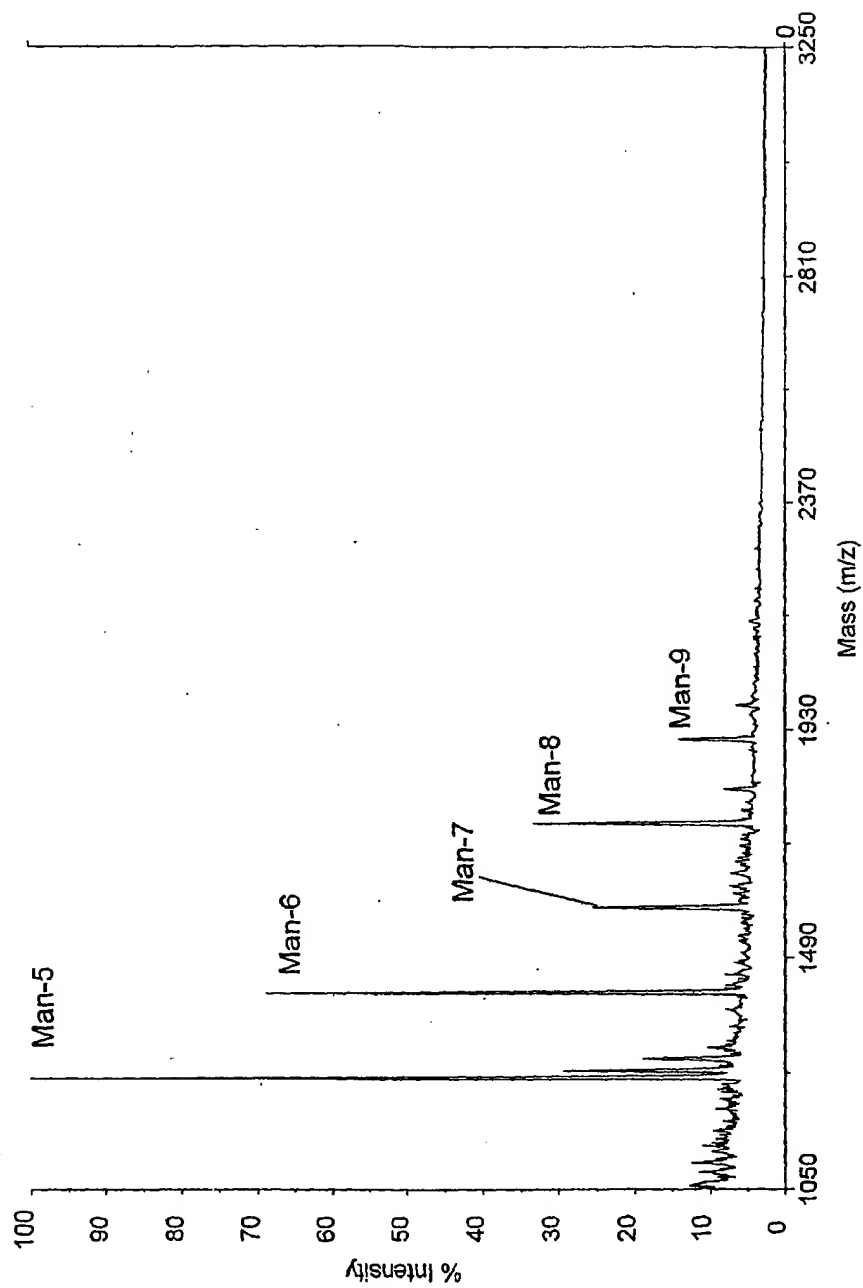


FIG. 5D

12/64

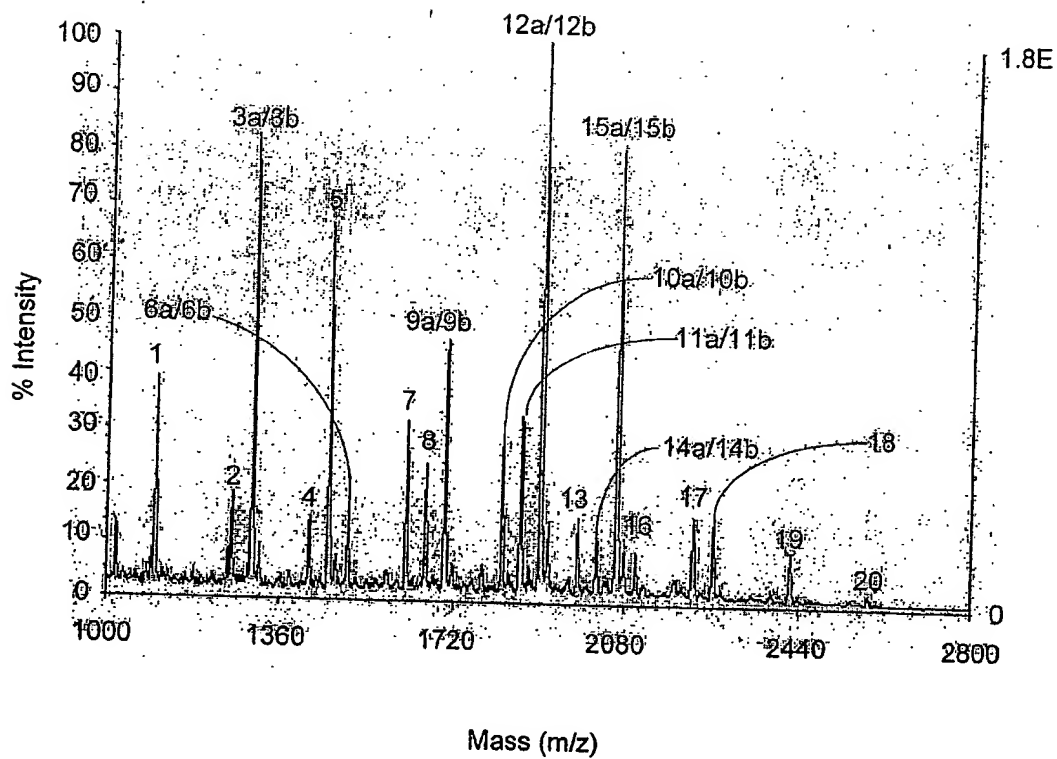


FIG. 6

13/64

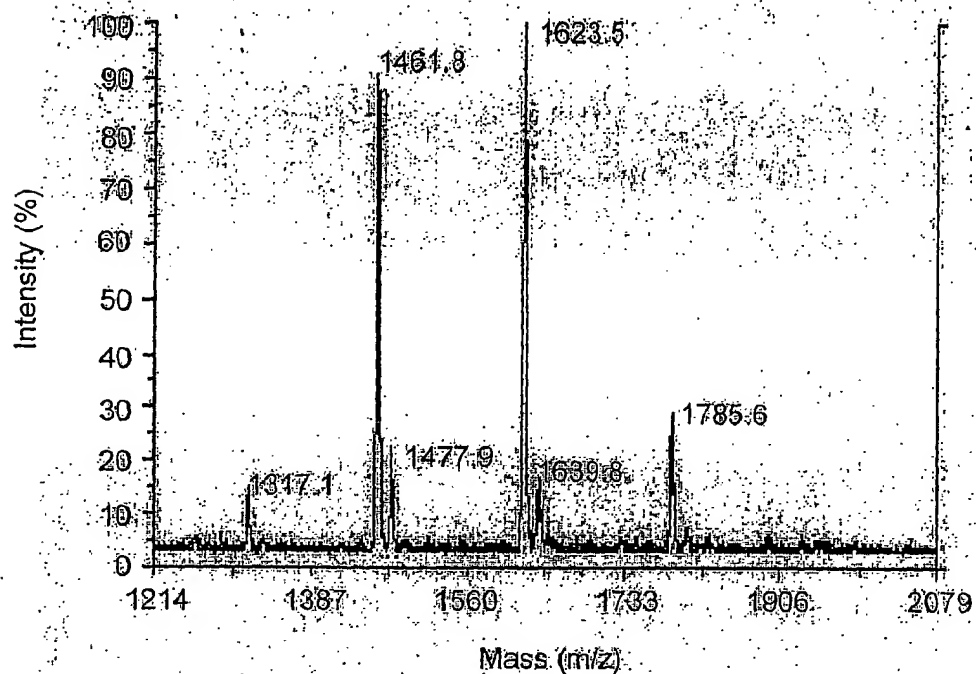


FIG. 7A

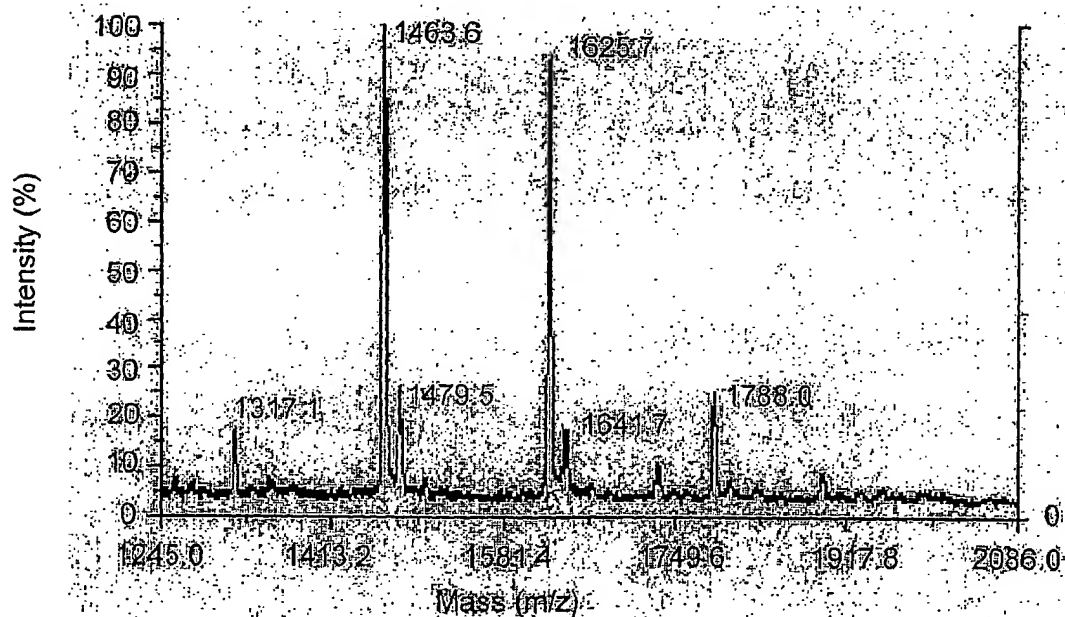


FIG. 7B

14/64

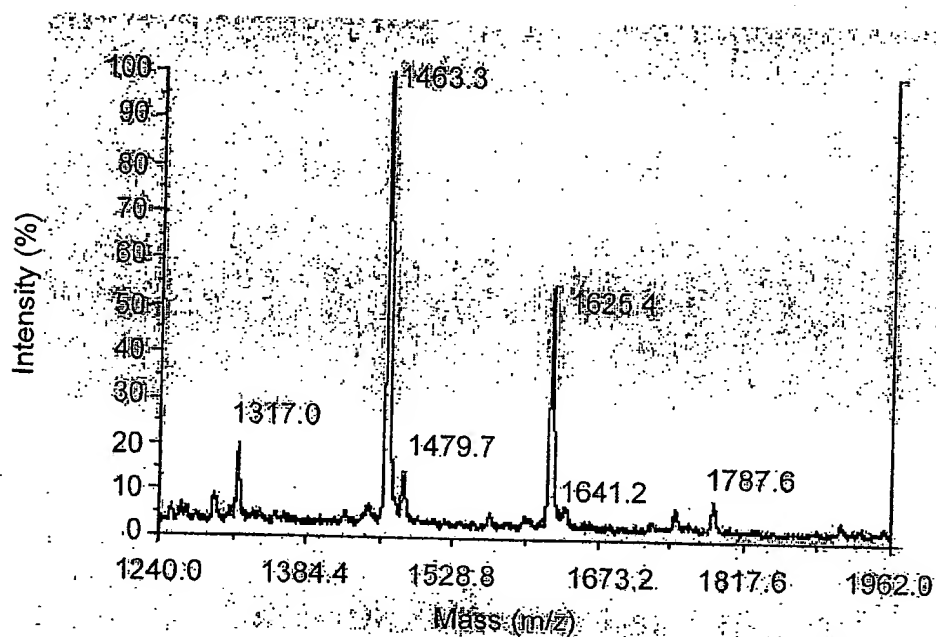


FIG. 7C

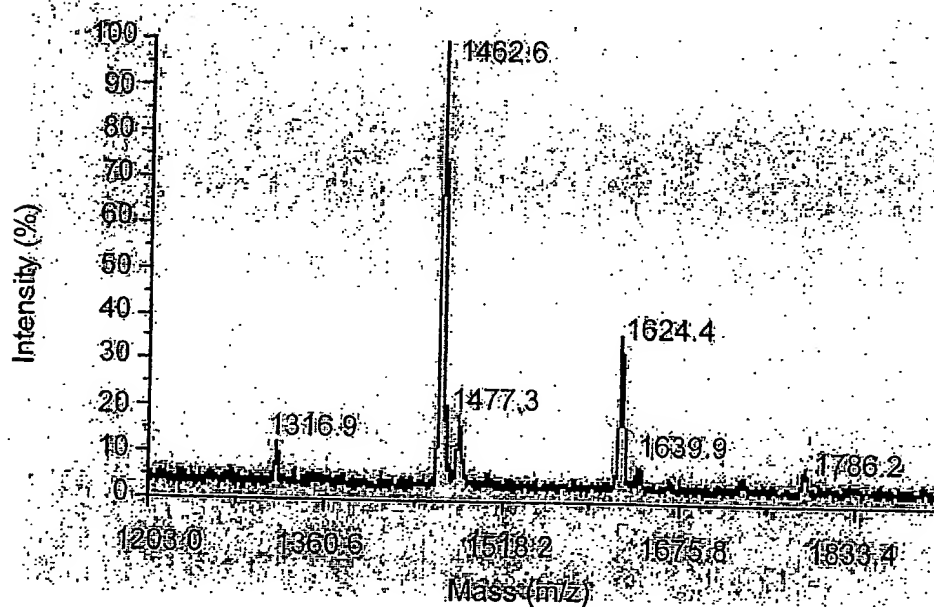


FIG. 7D

15/64

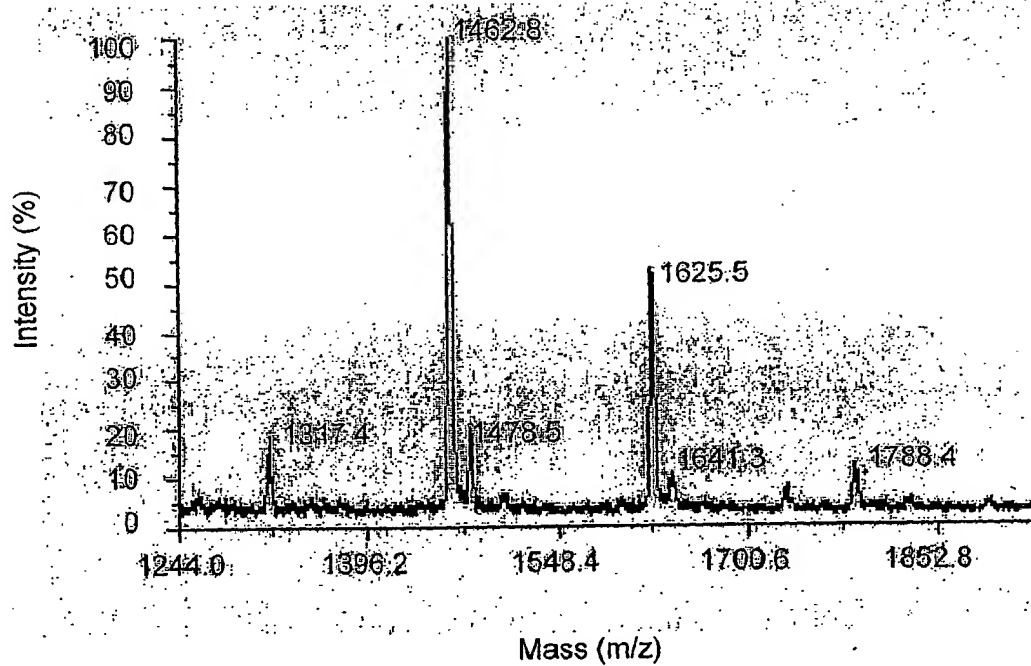


FIG. 7E

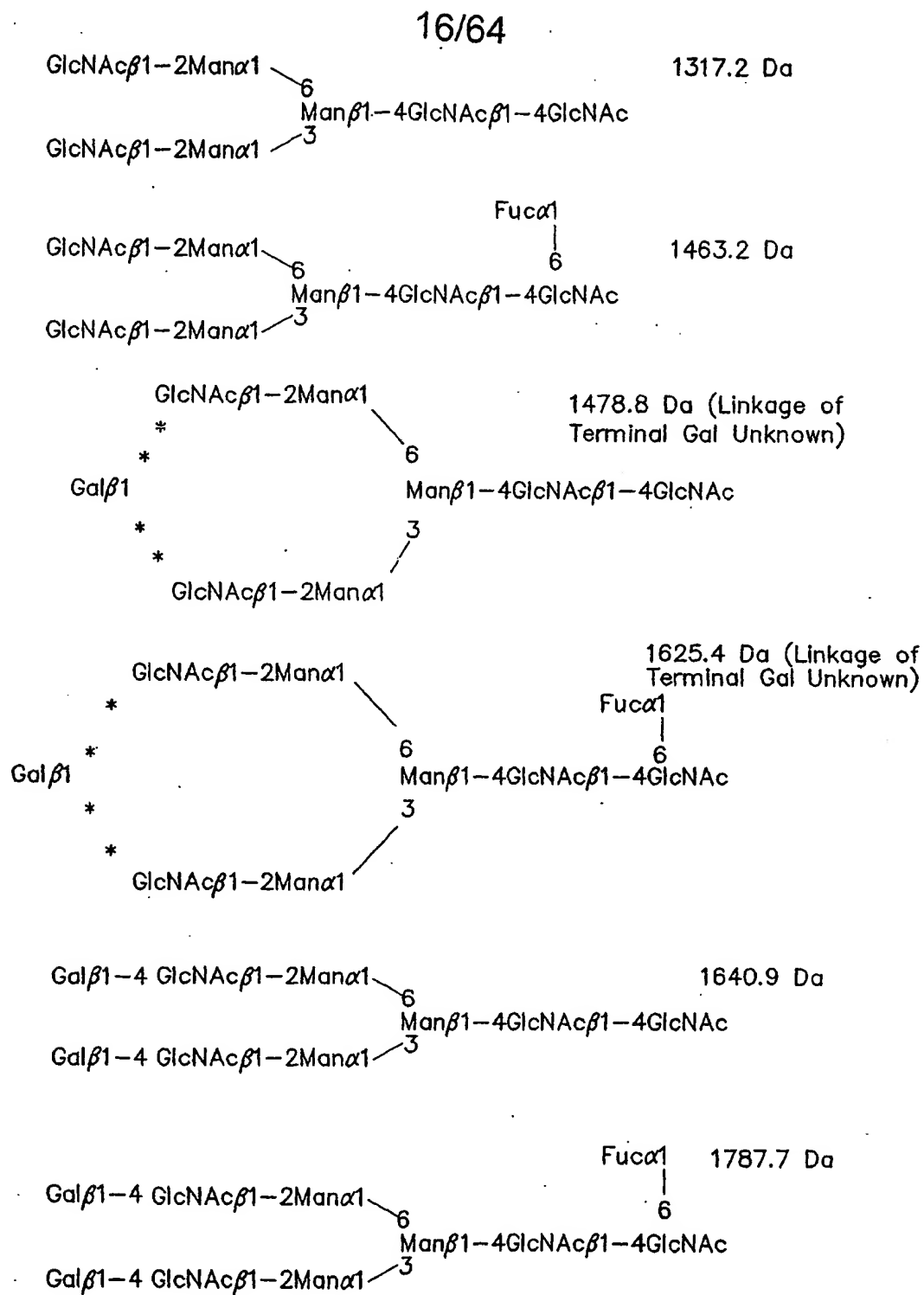


FIG. 8

17/64

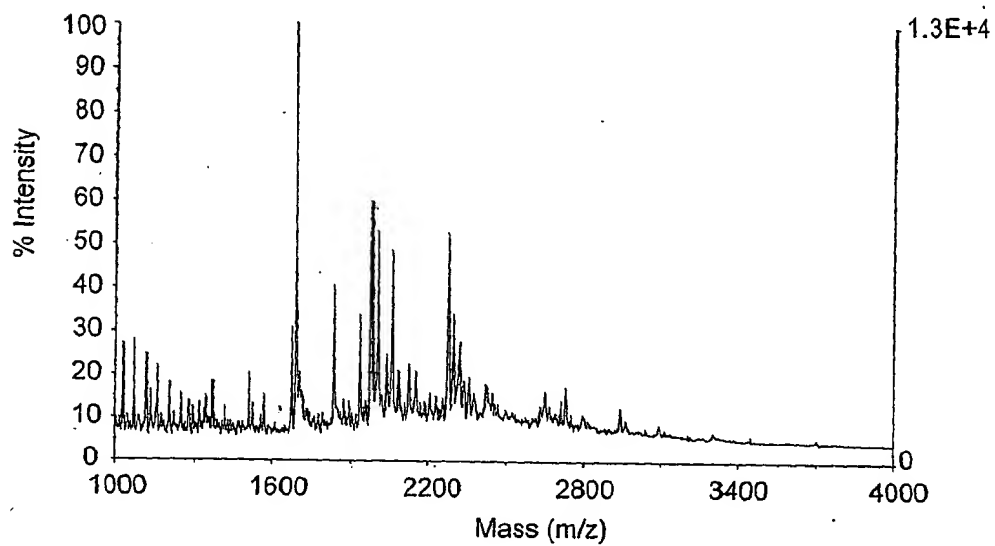


FIG. 9A

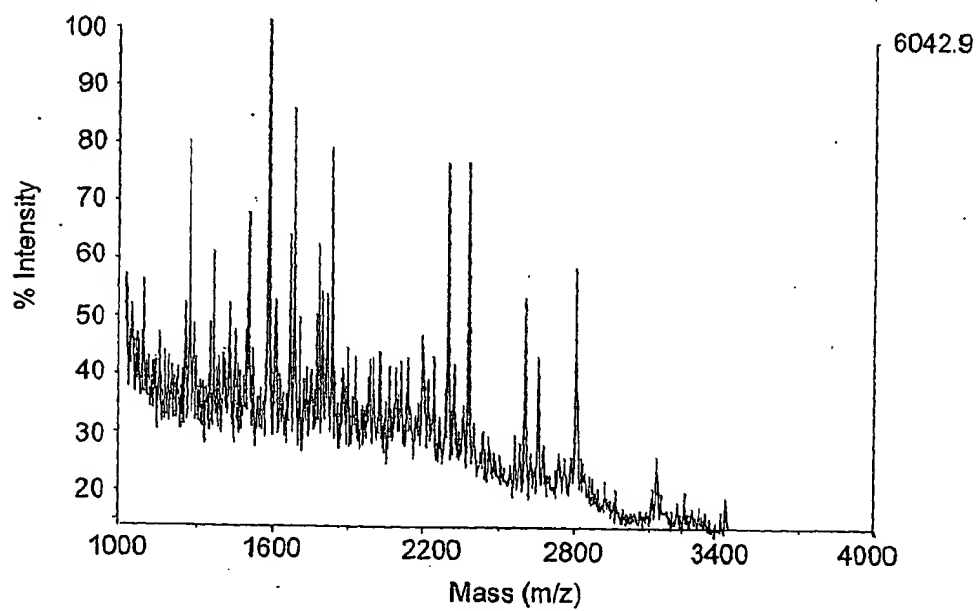


FIG. 9B

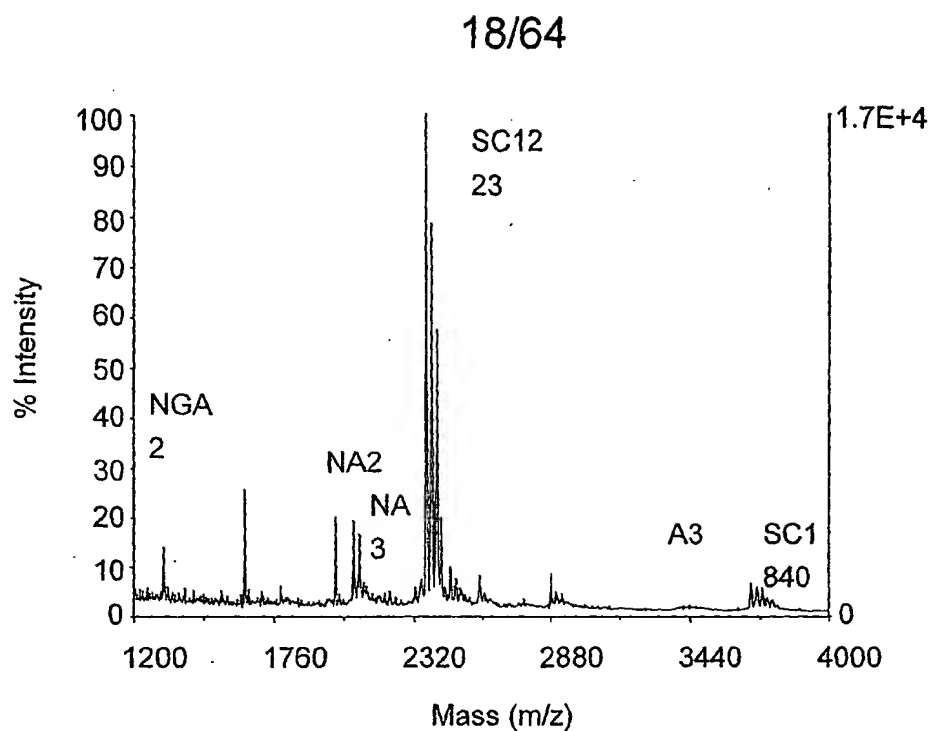


FIG. 10A

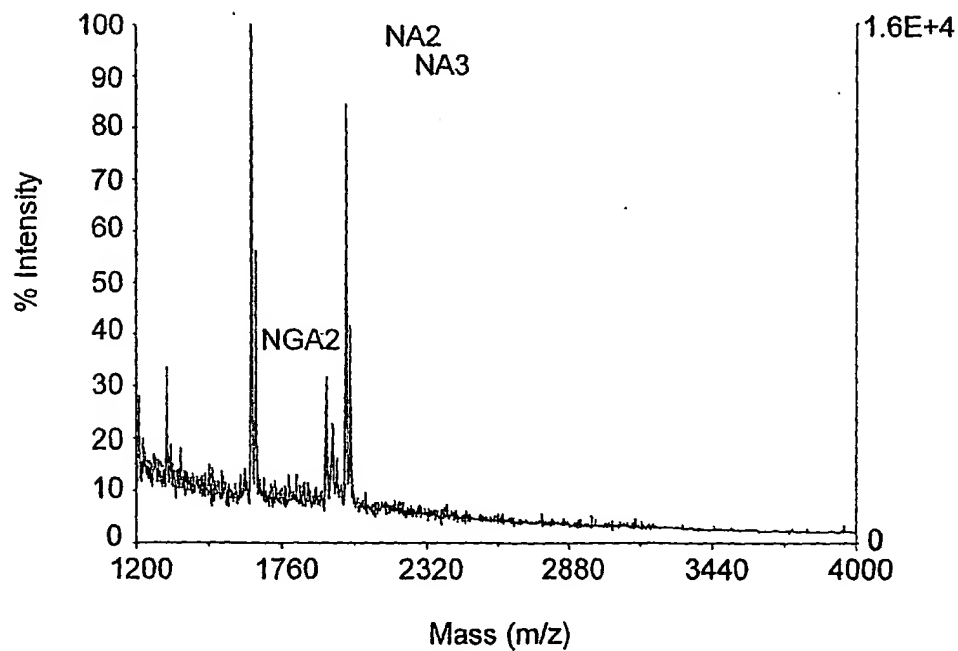


FIG. 10B

19/64

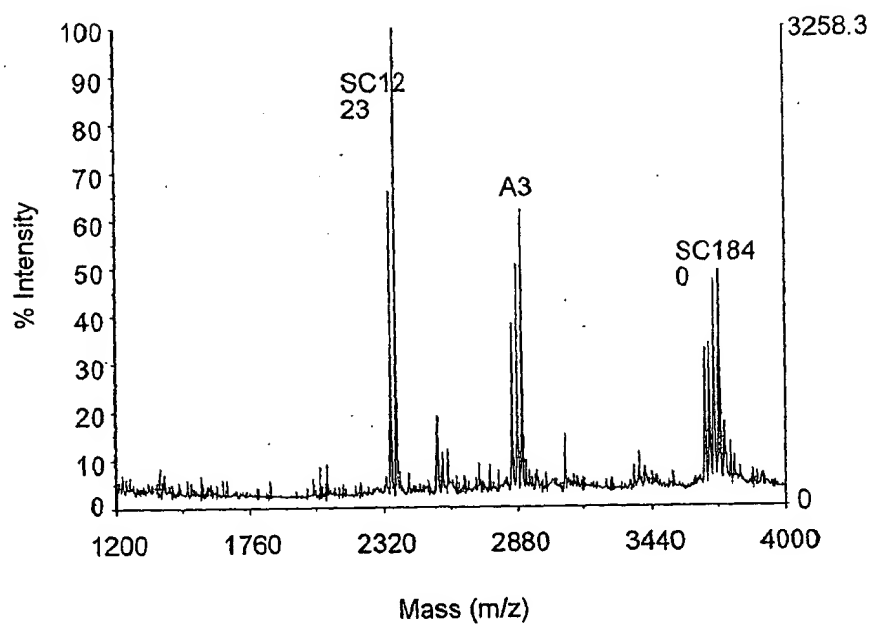


FIG. 10C

20/64

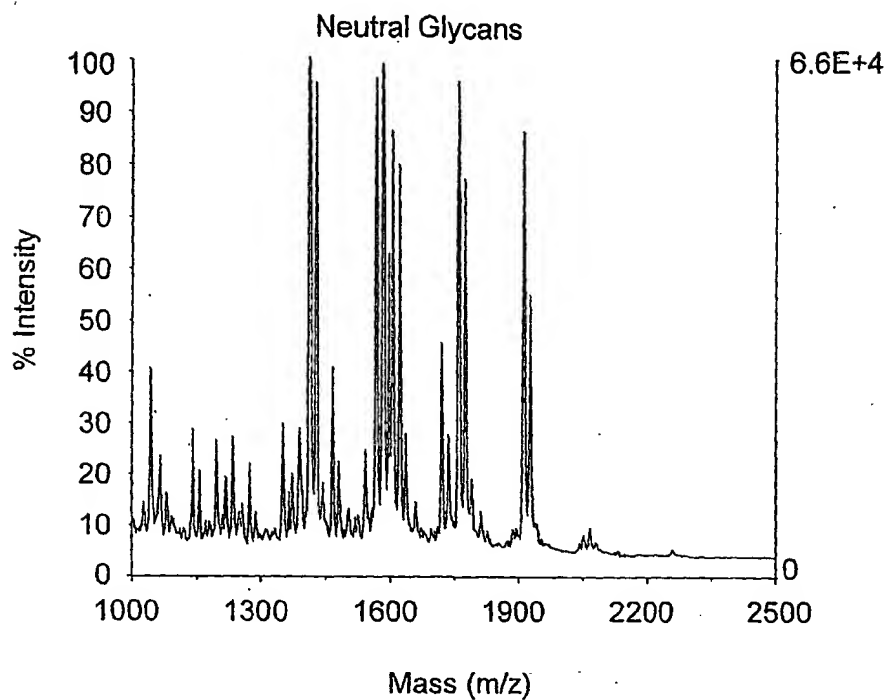


FIG. 11A

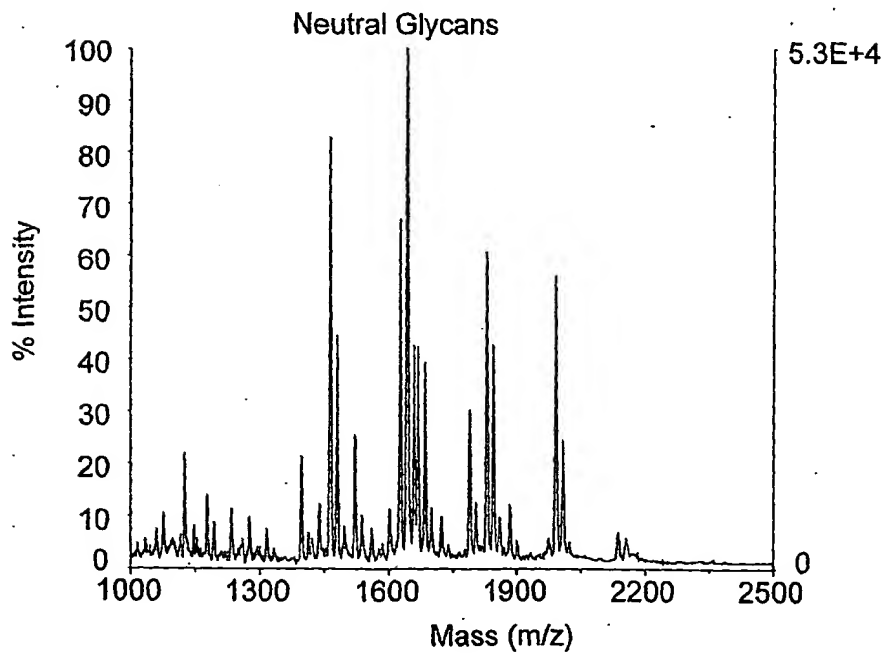


FIG. 11B

21/64

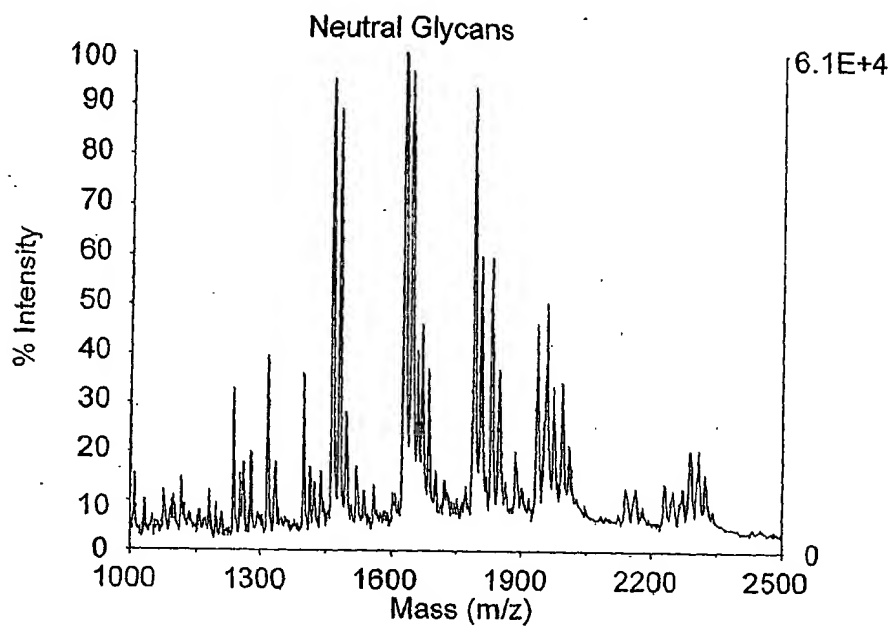


FIG. 11C

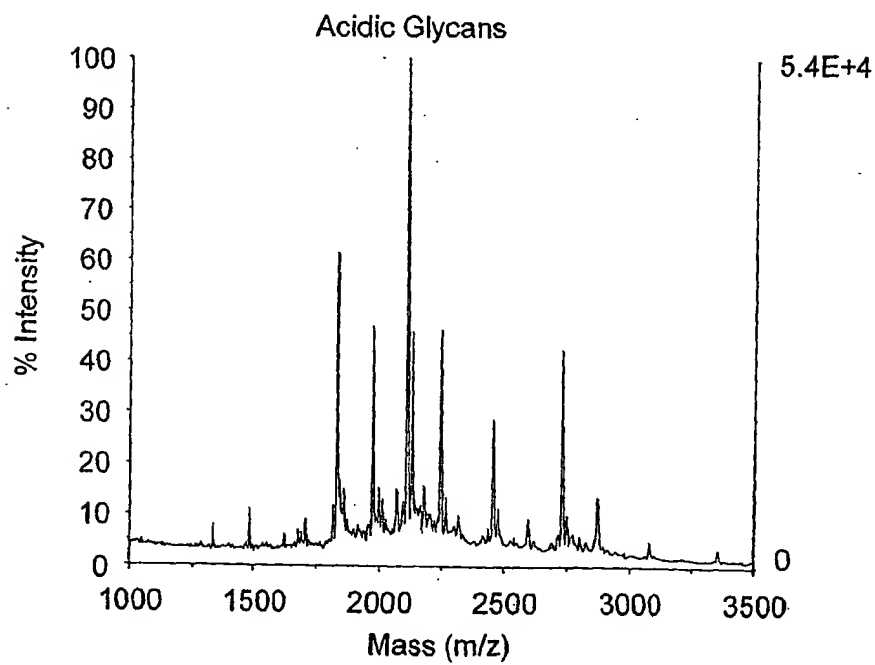
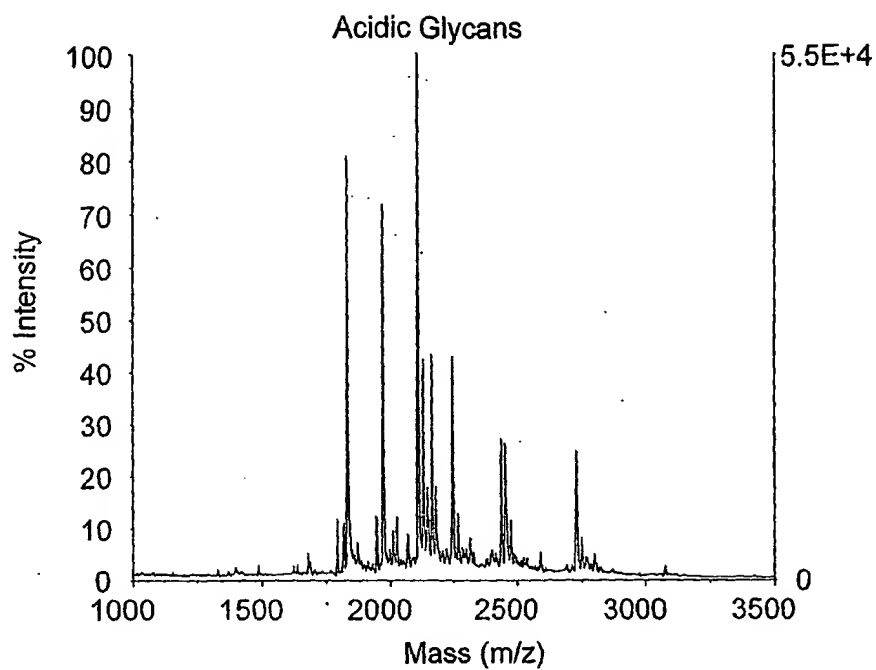
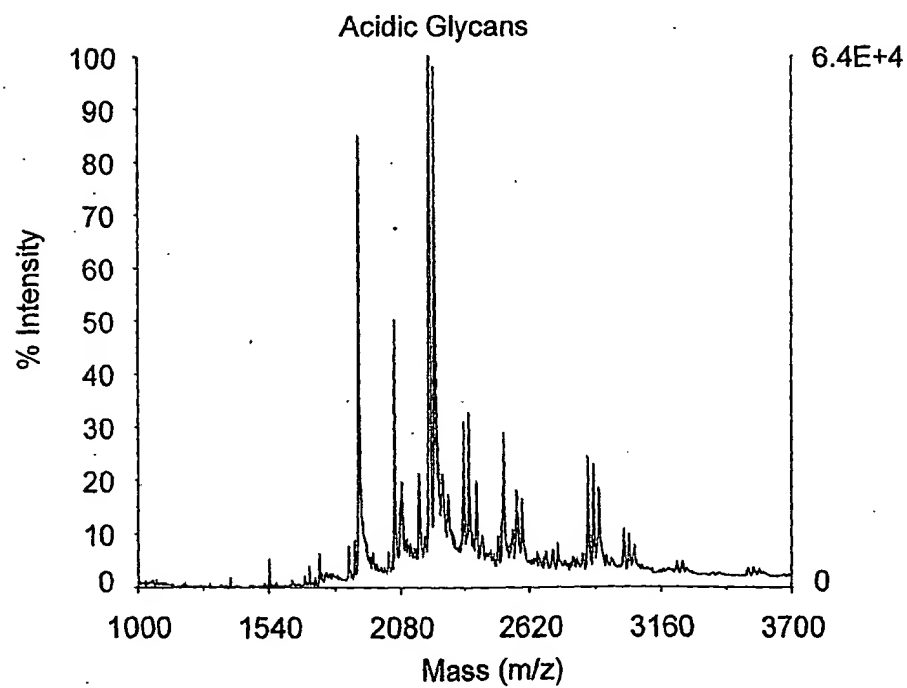


FIG. 11D

22/64

*FIG. 11E**FIG. 11F*

23/64

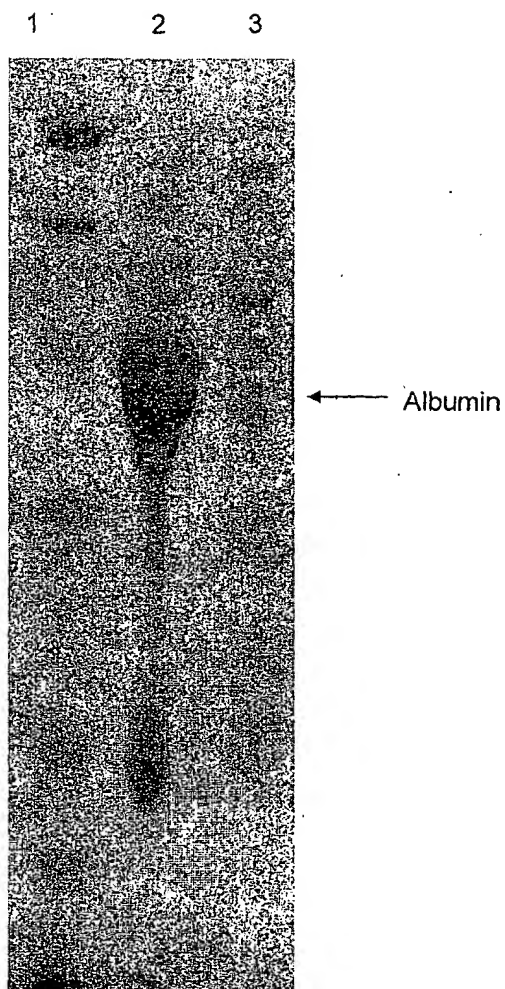


FIG. 12A

24/64

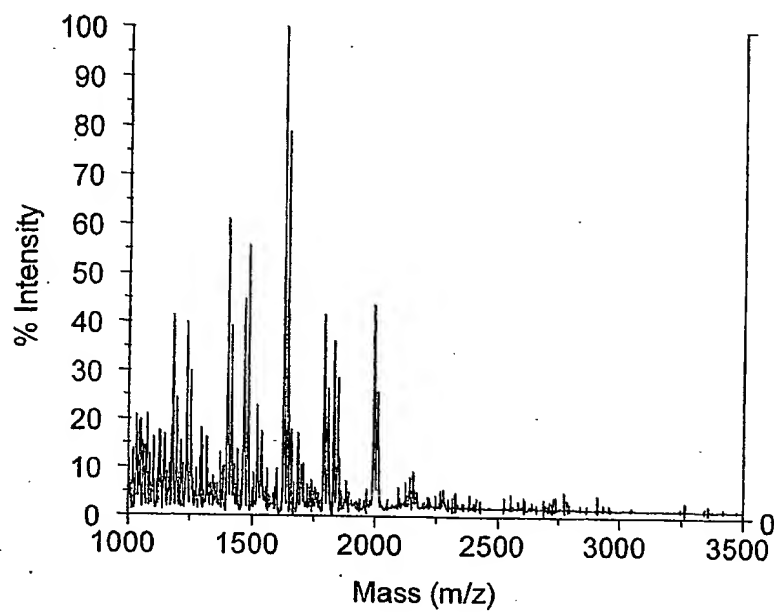


FIG. 12B

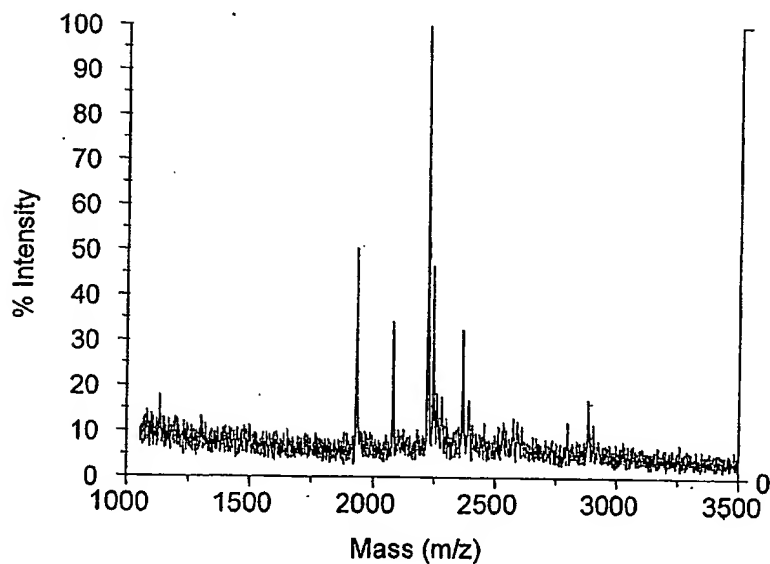


FIG. 12C

25/64

1 2 3 4



FIG. 13A

26/64

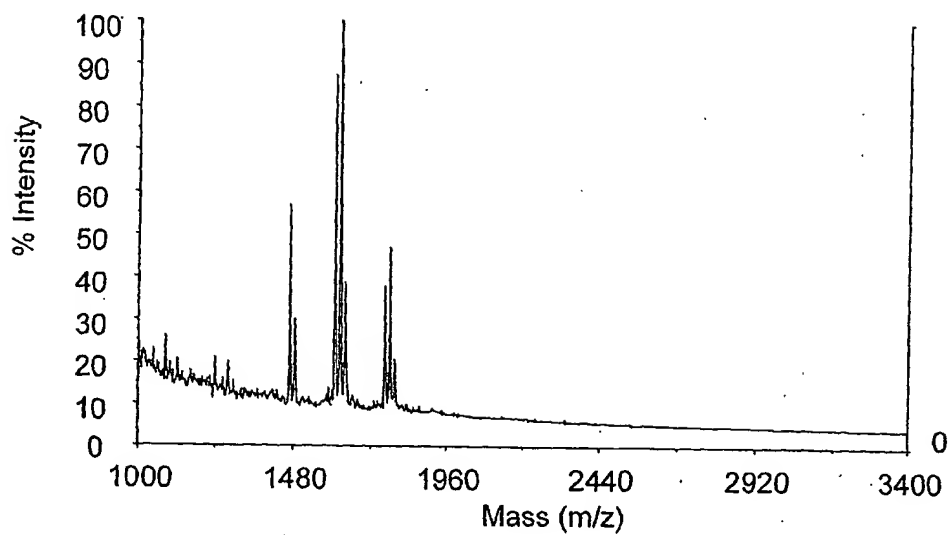


FIG. 13B

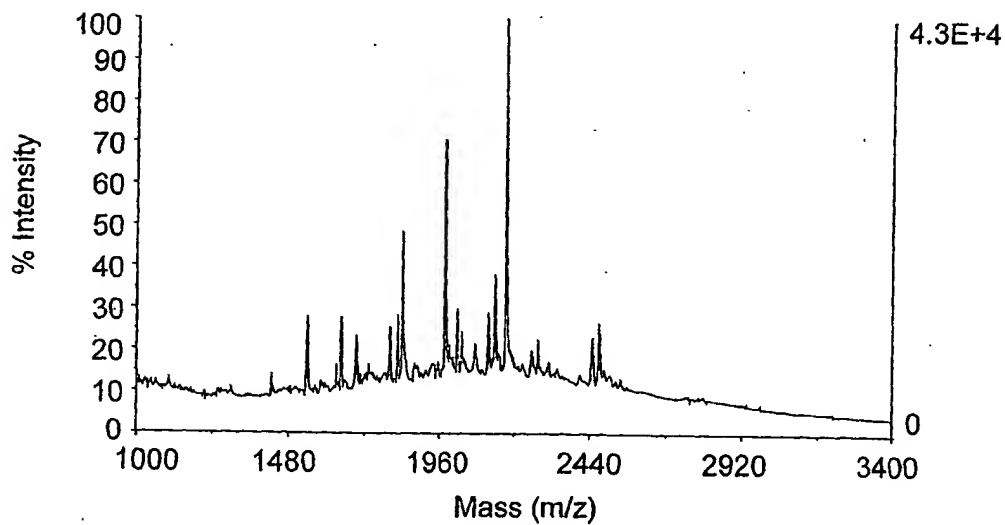


FIG. 13C

27/64

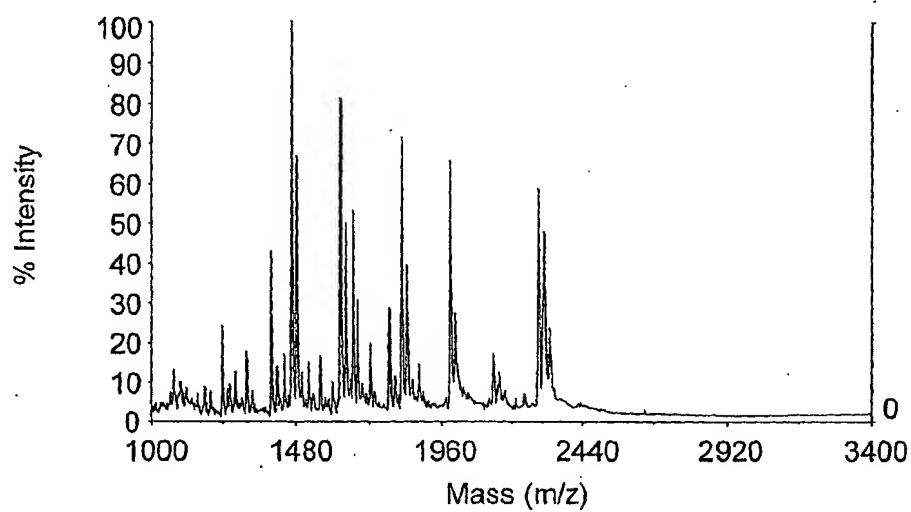


FIG. 13D

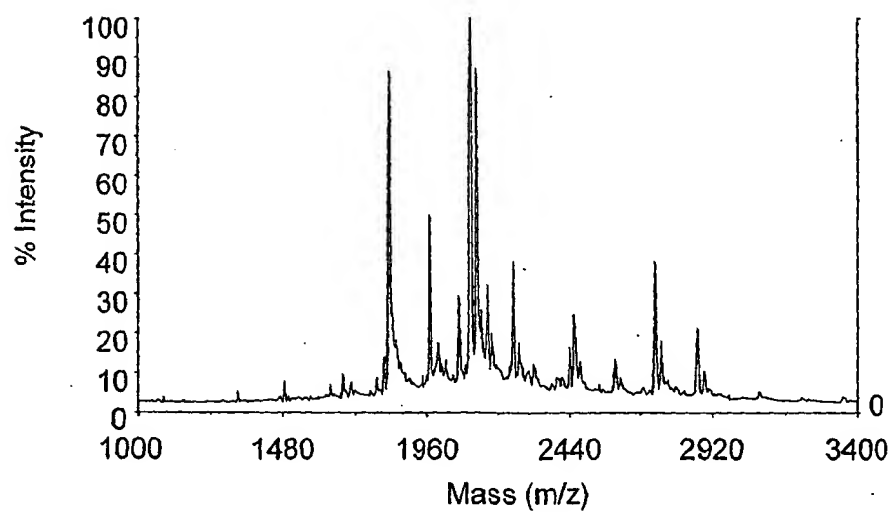


FIG. 13E

28/64

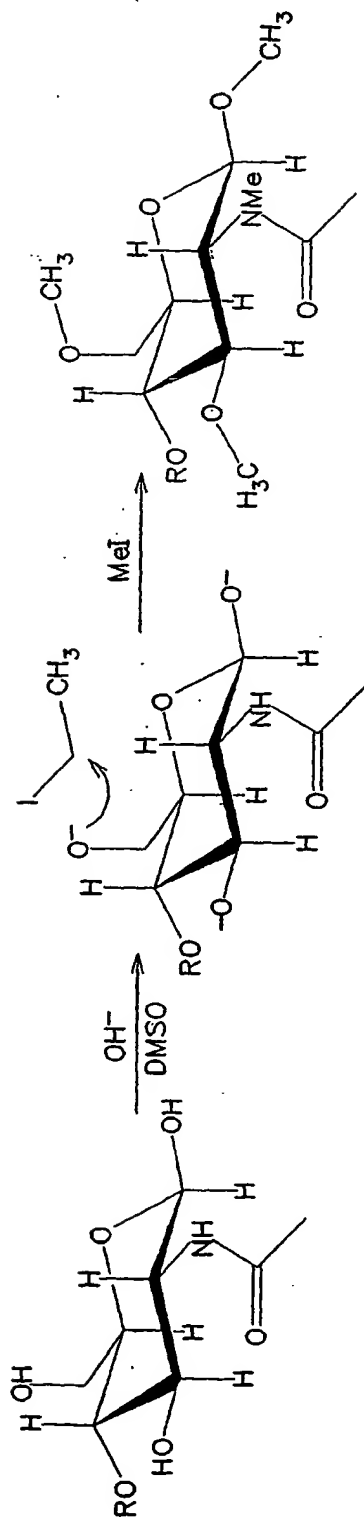


FIG. 14

29/64

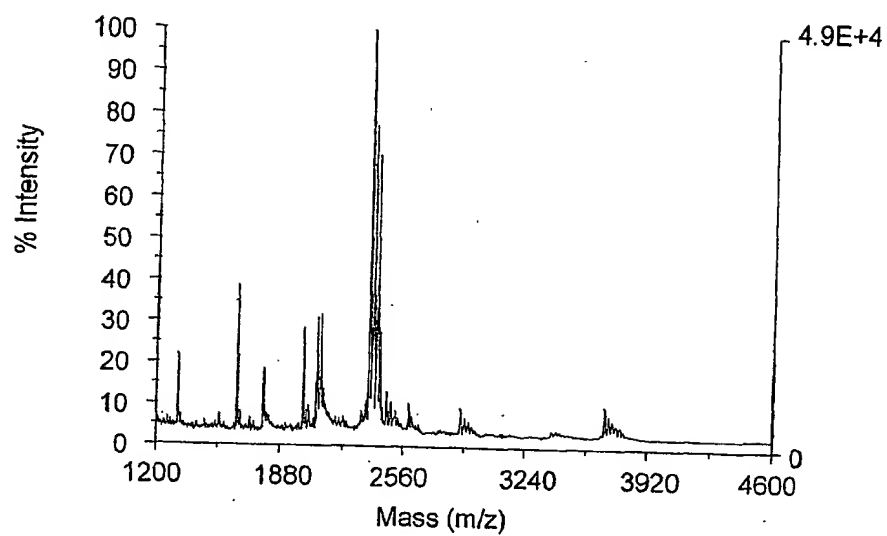


FIG. 15A

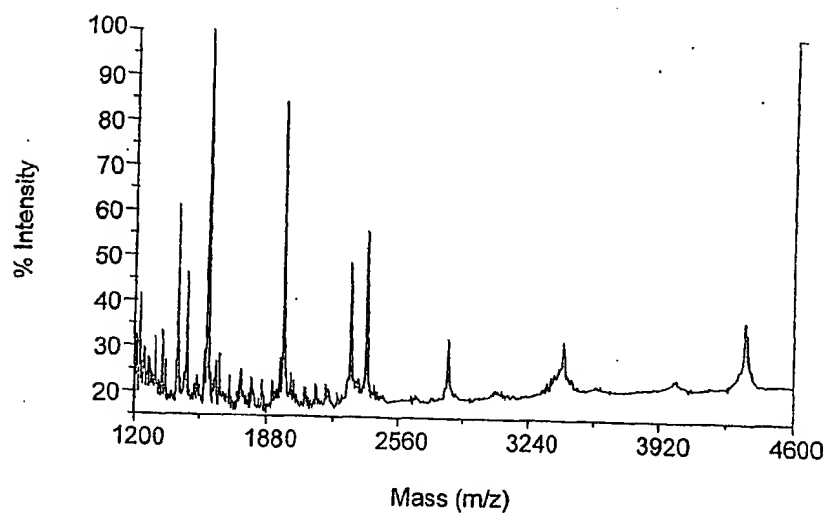


FIG. 15B

30/64

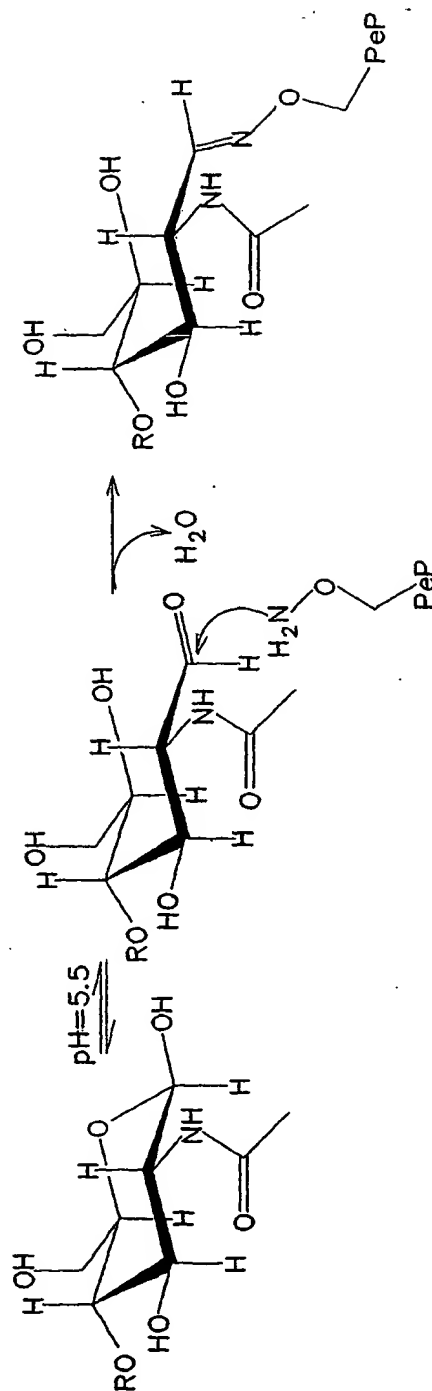
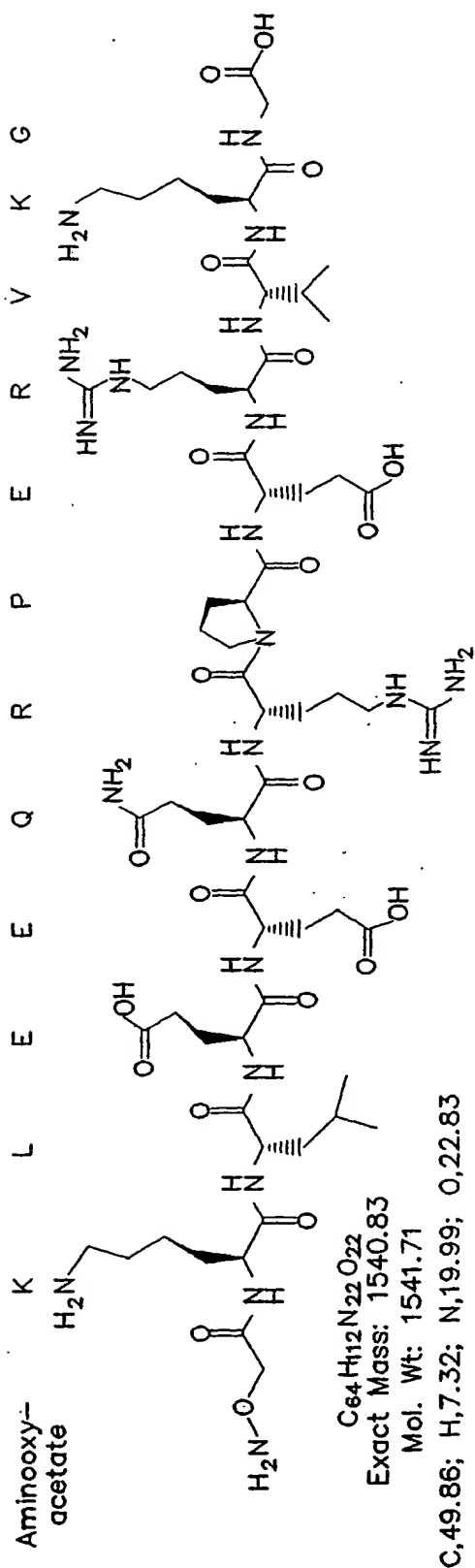


FIG. 16

31/64

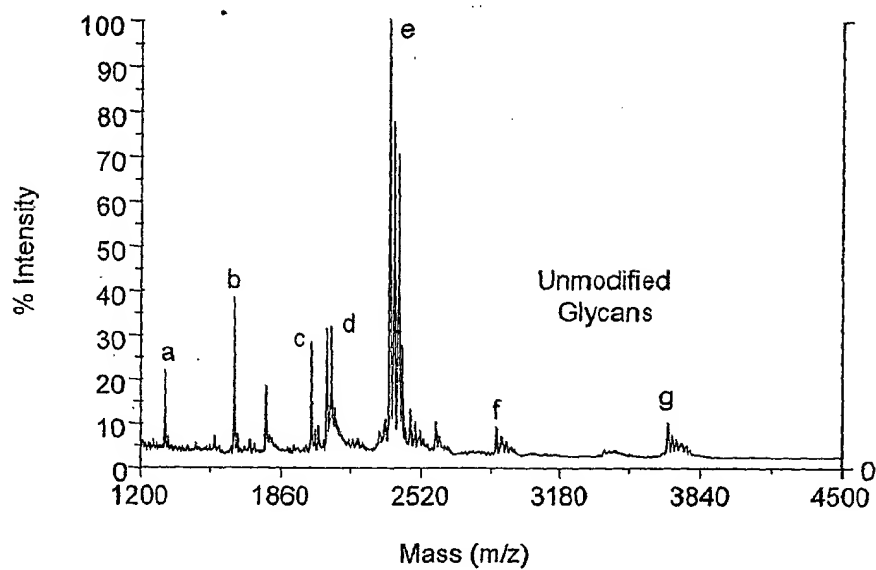


FIG. 17A

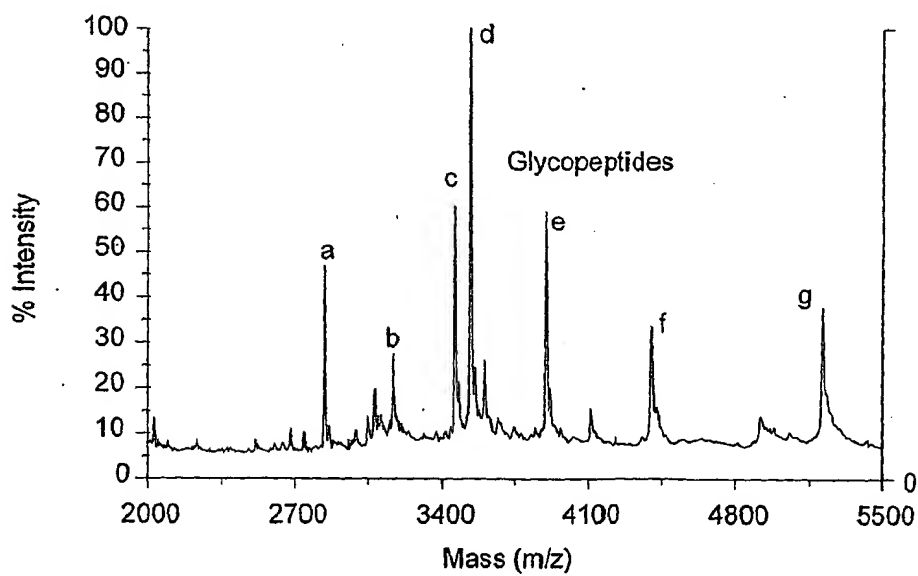


FIG. 17B

32/64

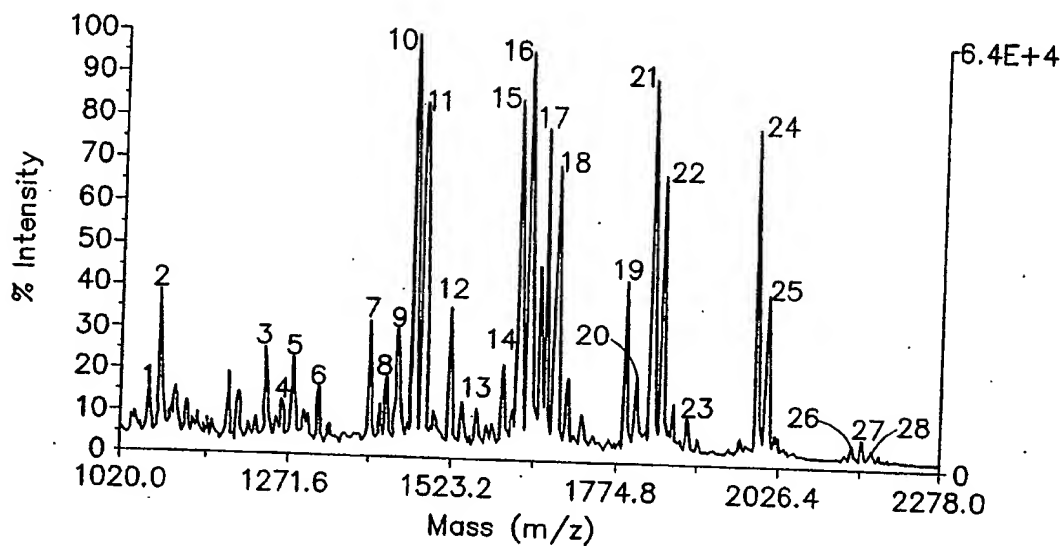


FIG. 18A

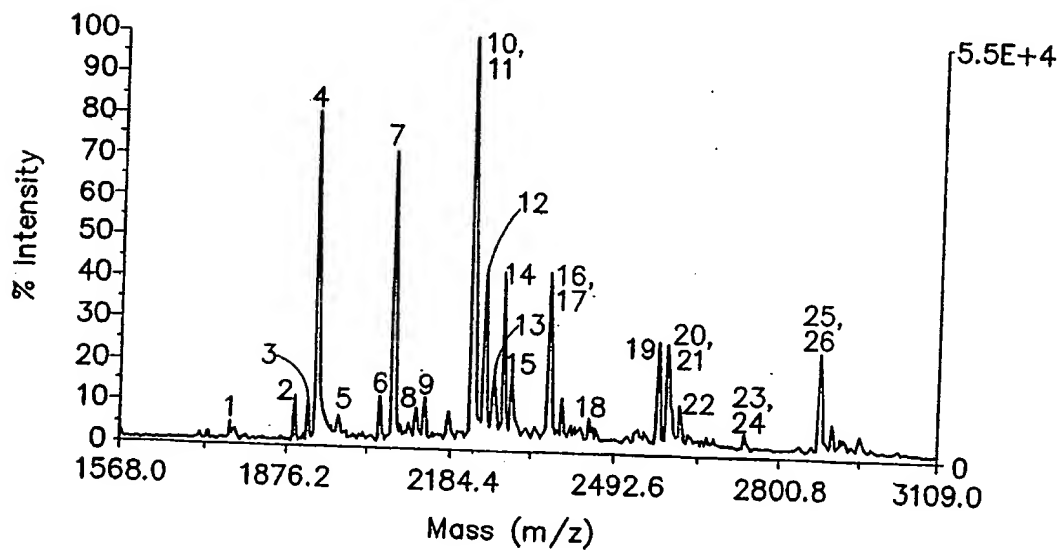


FIG. 18B

33/64

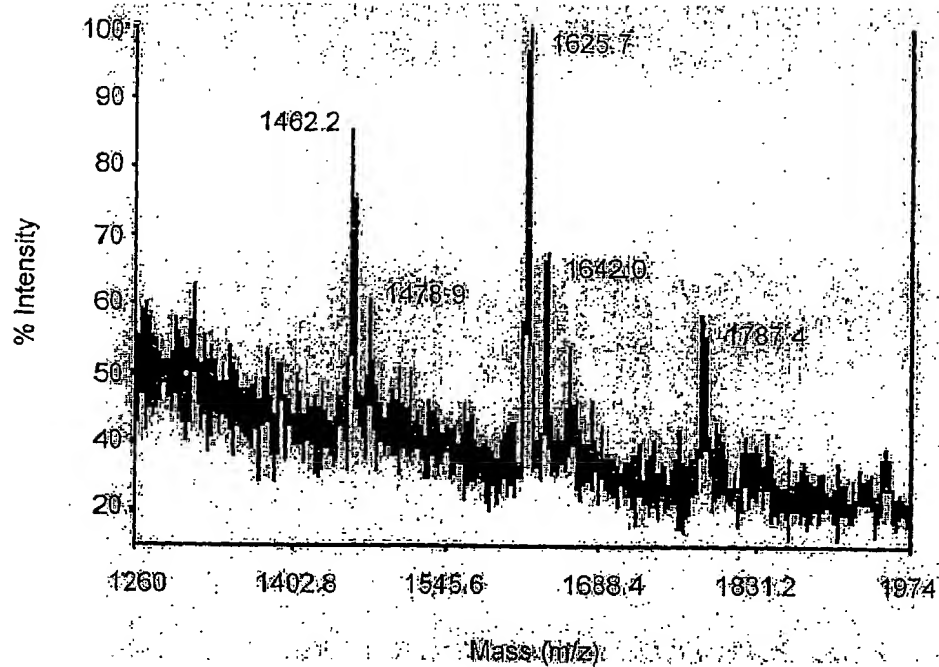
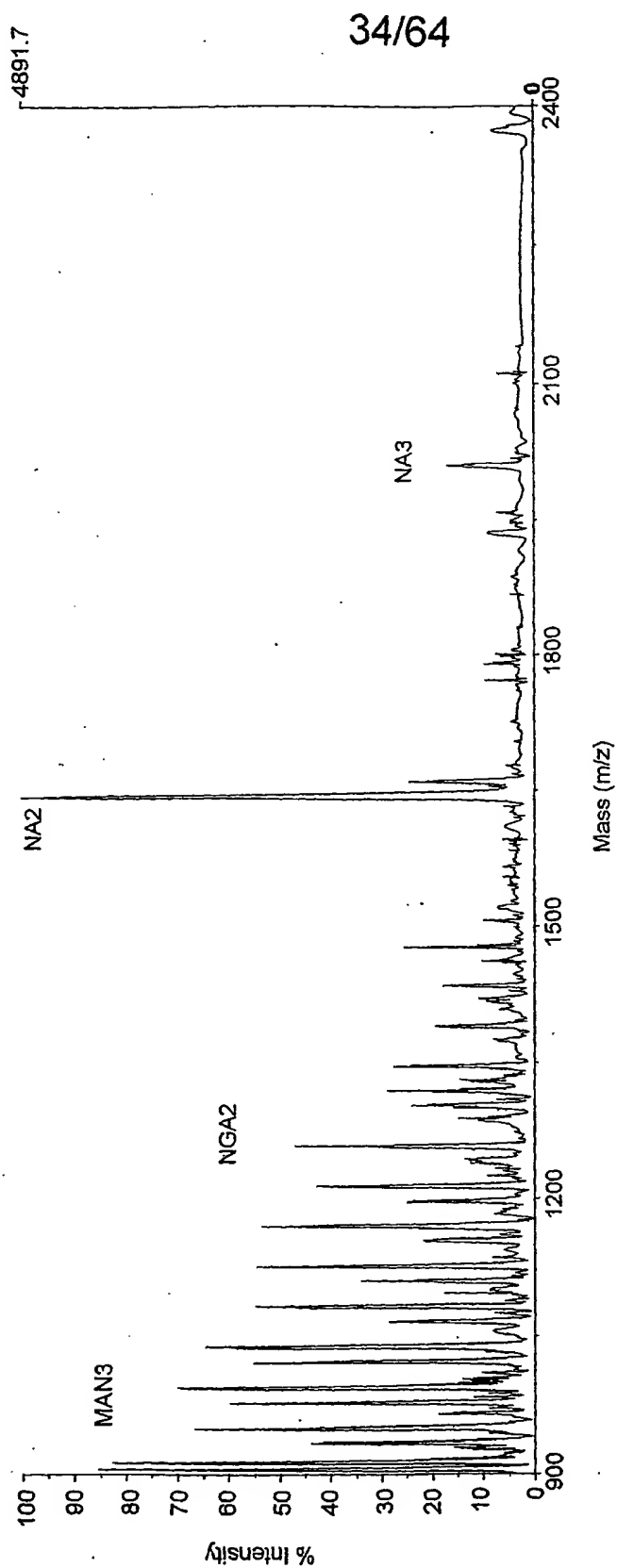


FIG. 19



DHB-Spermine Neutrals, 1 pmol

FIG. 20A

35/64

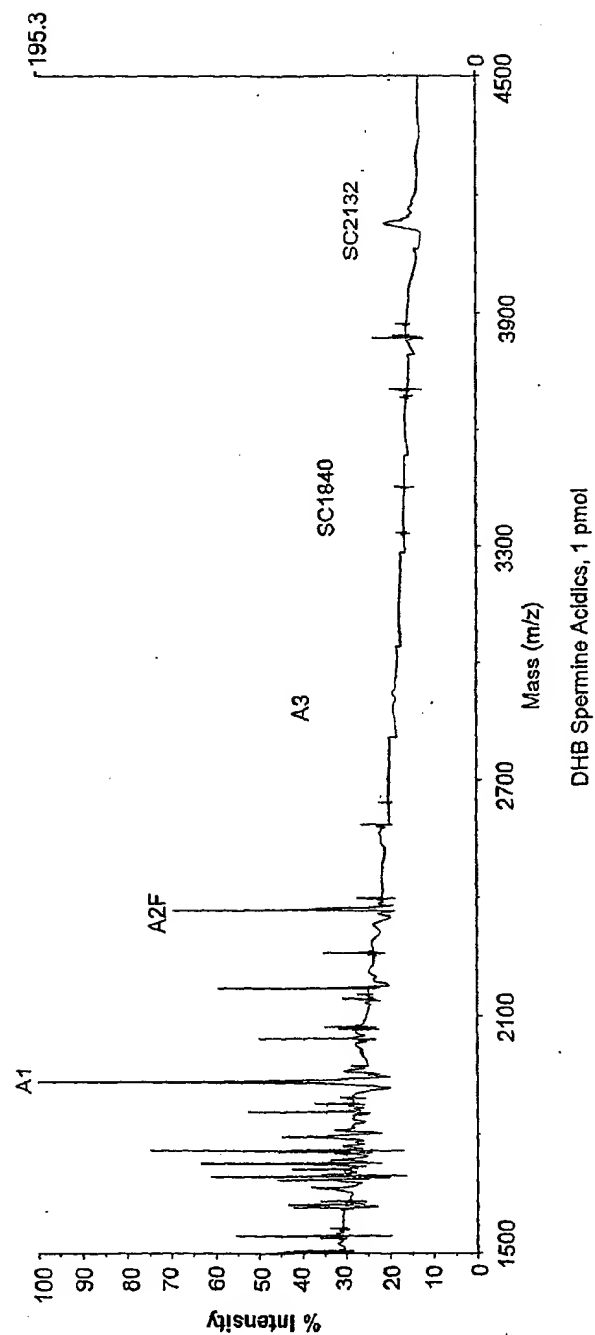
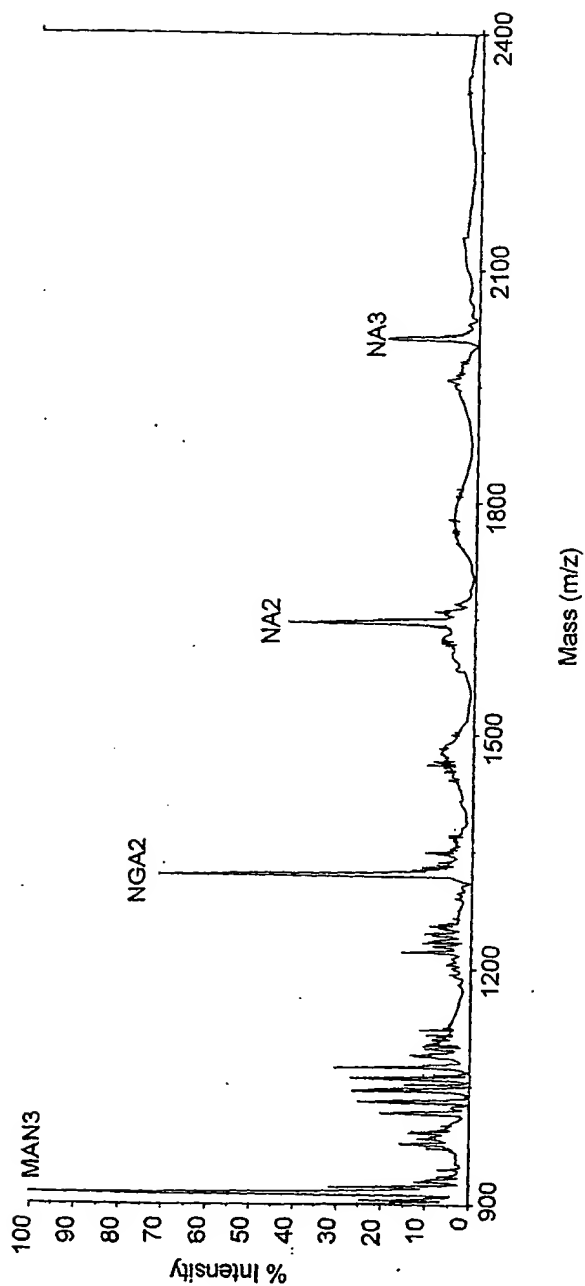


FIG. 20B

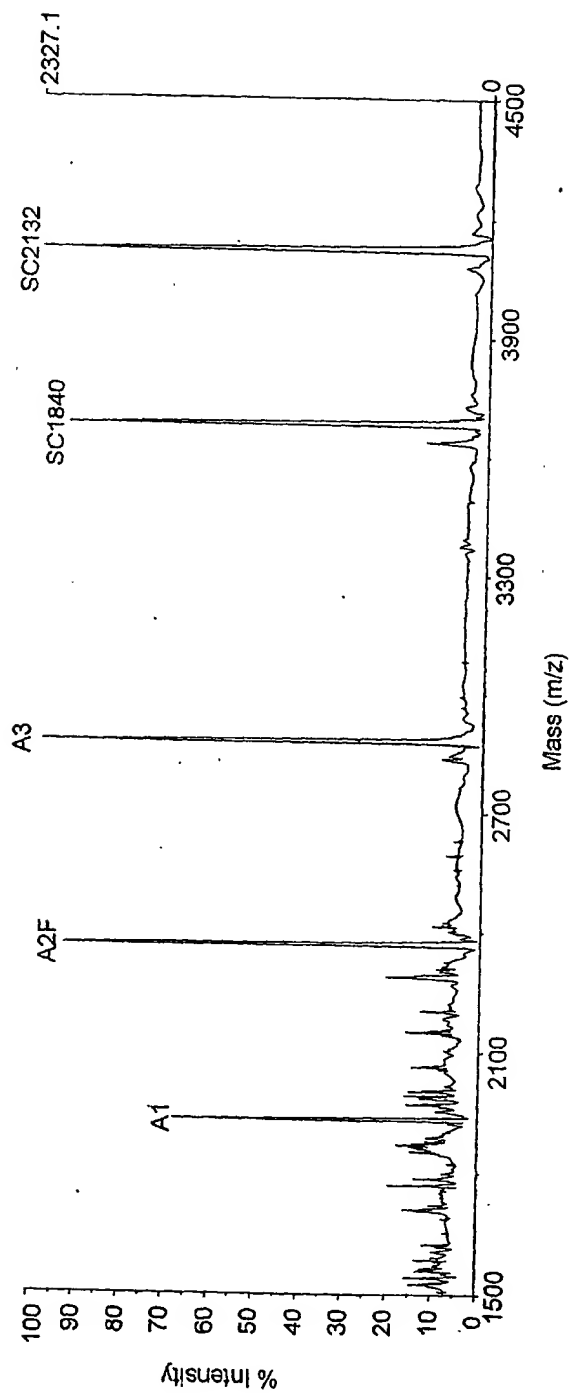
36/64



5-MSA/DHB Neutrals, 25 fmol

FIG. 20C

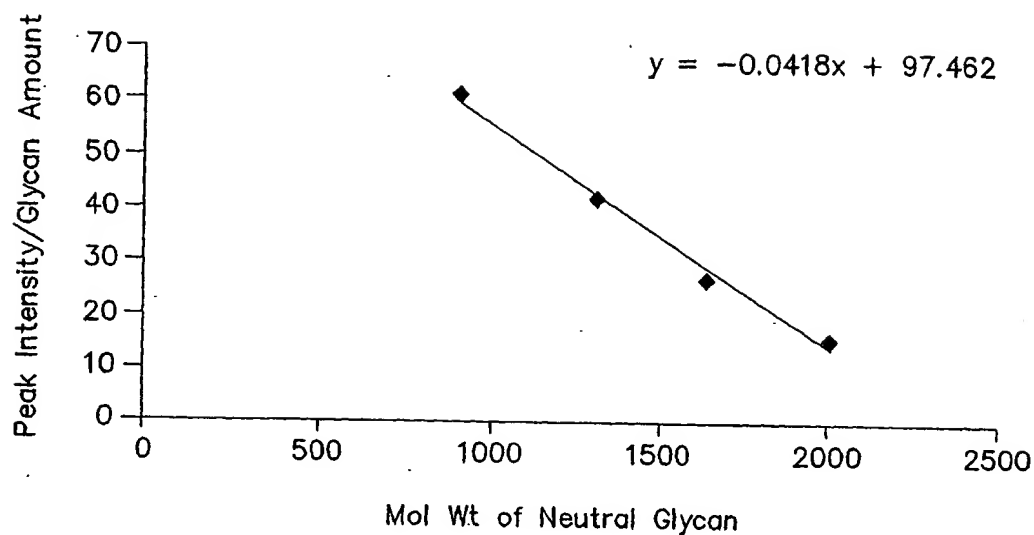
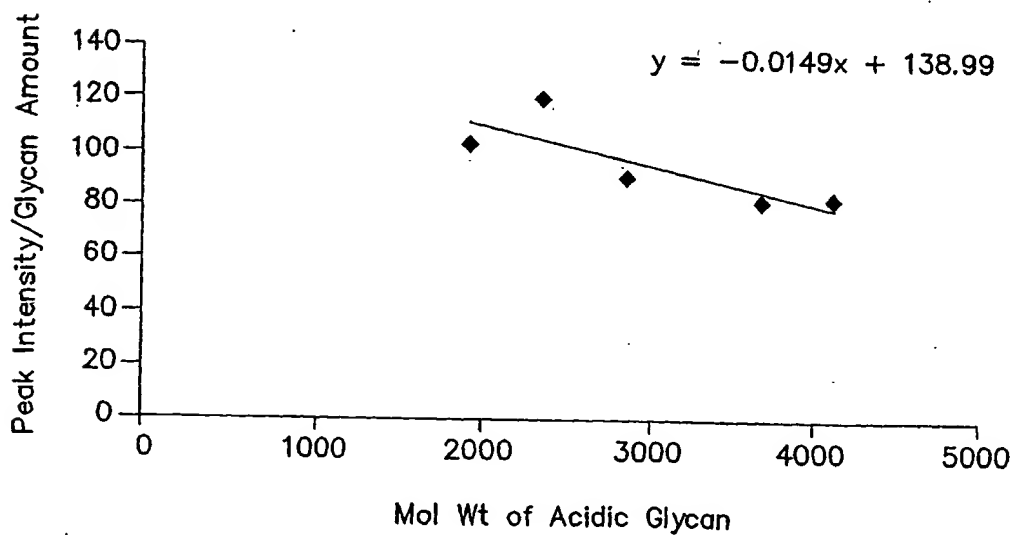
37/64



Nafion + ATT Acids, 25 fmol

FIG. 20D

38/64

**FIG. 21A****FIG. 21B**

39/64

Glycans in Spot	Man3 Intens.	NGA2 Intens.	NA2 Intens.	NGA3 Intens.
Man3	5.9e3	—	—	—
Man3 NGA2	6.5e3	4.4e3	—	—
Man3 NGA2 NA2	6.4e3	4.5e3	2.0e3	—
Man3 NGA2 NA2	8.4e3	4.8e3	2.3e3	—

FIG. 21C

Glycans in Spot	A1 Intens.	A2F Intens.	A3 Intens.	SC1840 Intens.	SC2132 Intens.
A1	2.2e3	—	—	—	—
A1 A2F	4.6e3	2.3e3	—	—	—
A1 A2F A3	2.6e3	1.2e3	3.6e3	—	—
A1 A2F A3 SC1840	3.2e3	2.0e3	5.1e3	3.5e3	—
A1 A2F A3 SC1840 SC2132	2.8e3	1.6e3	3.9e3	2.0e3	2.5e3

FIG. 21D

40/64

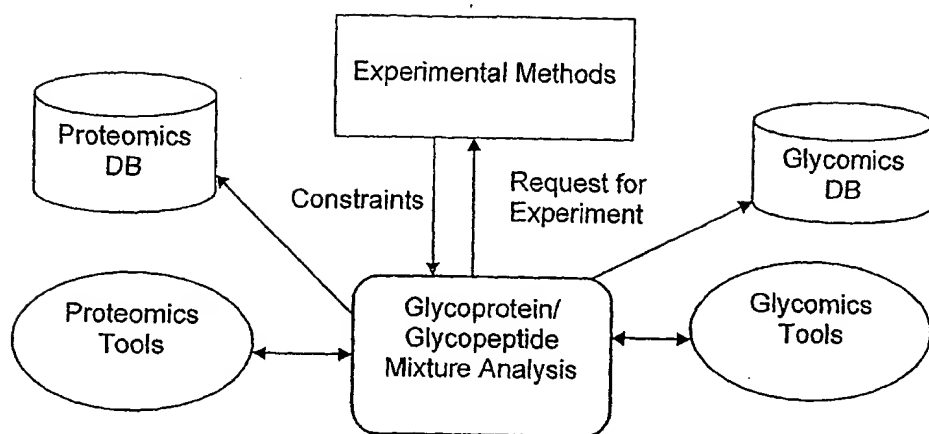


FIG. 22

41/64

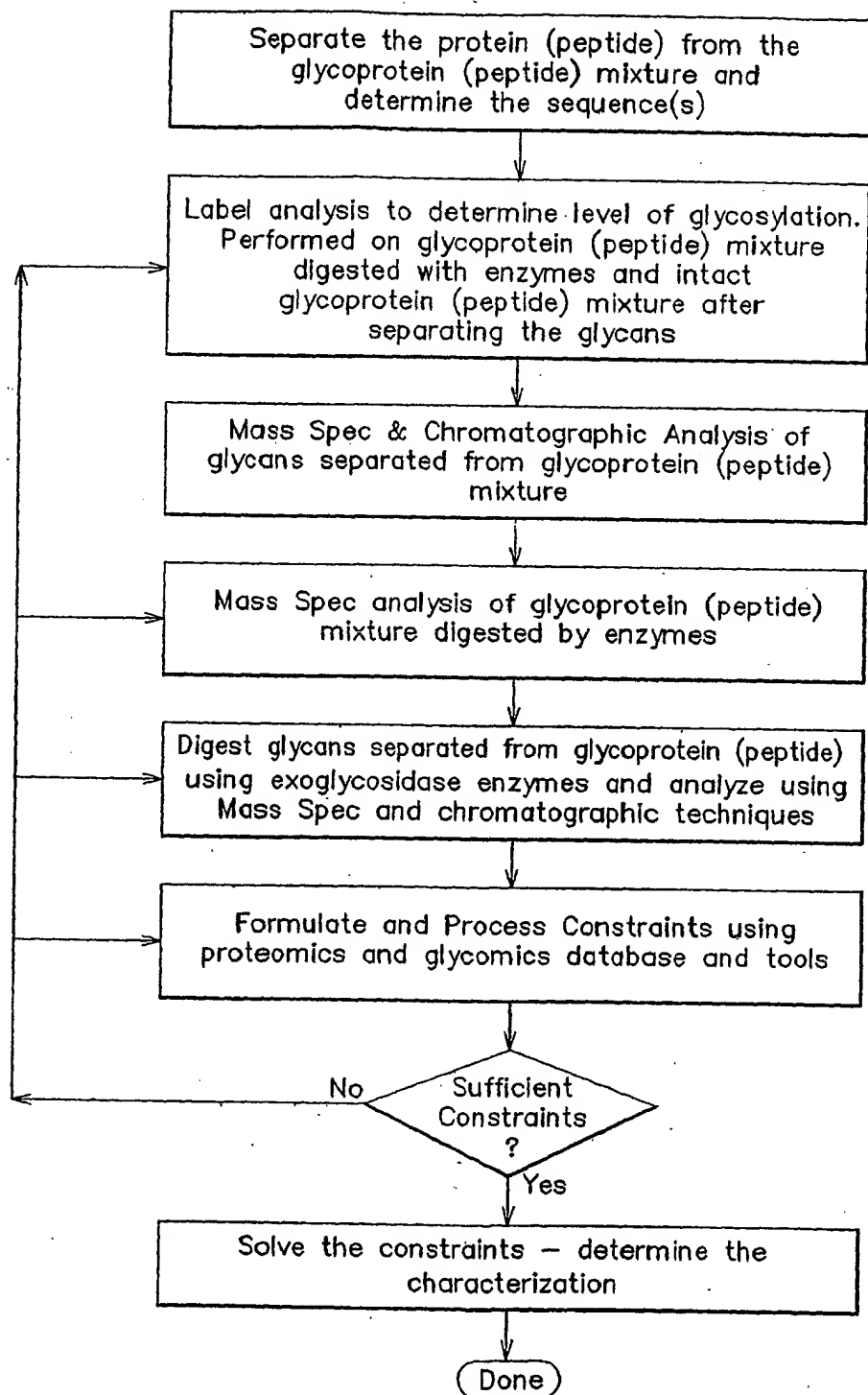


FIG. 23

42/64

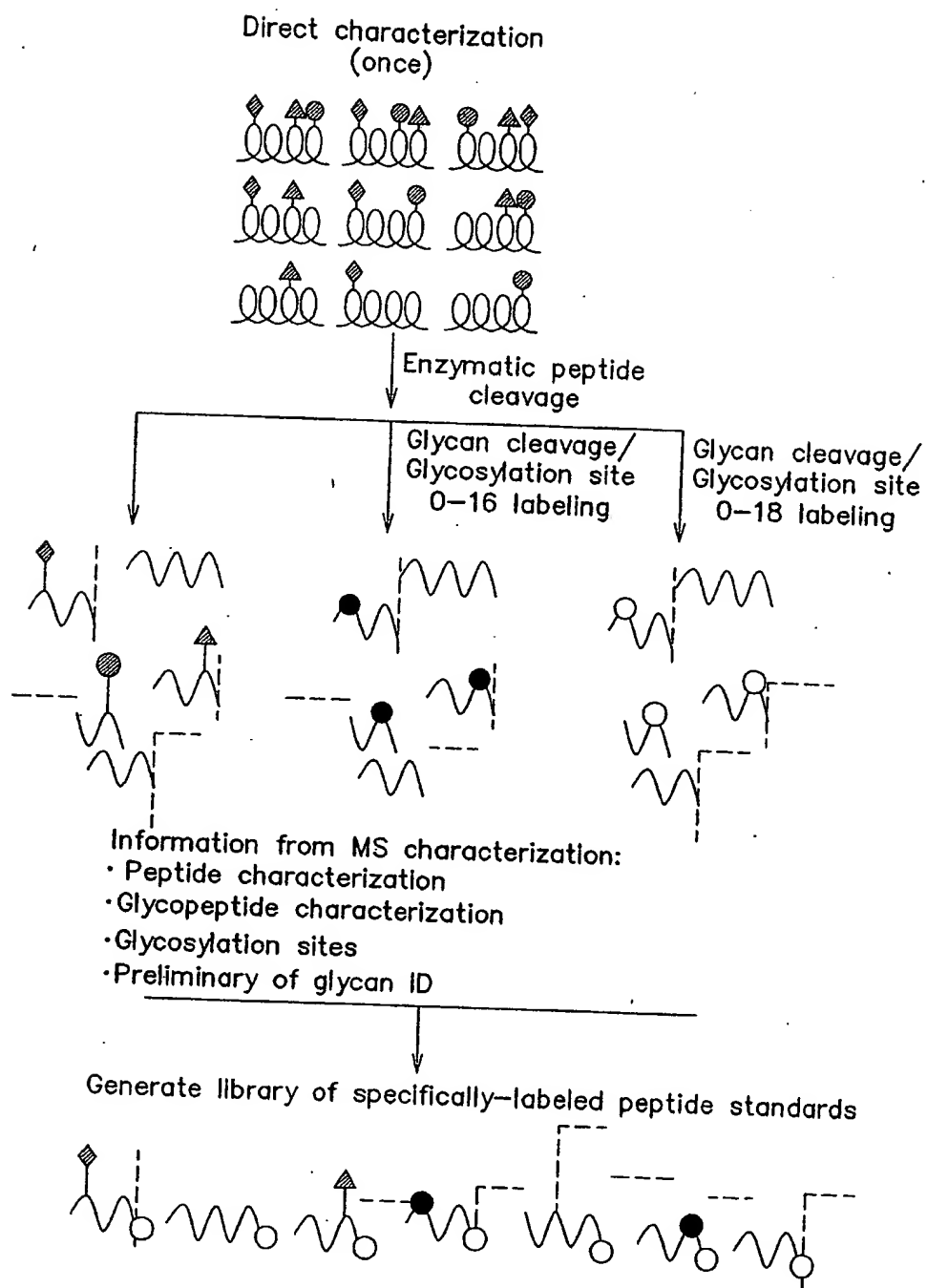
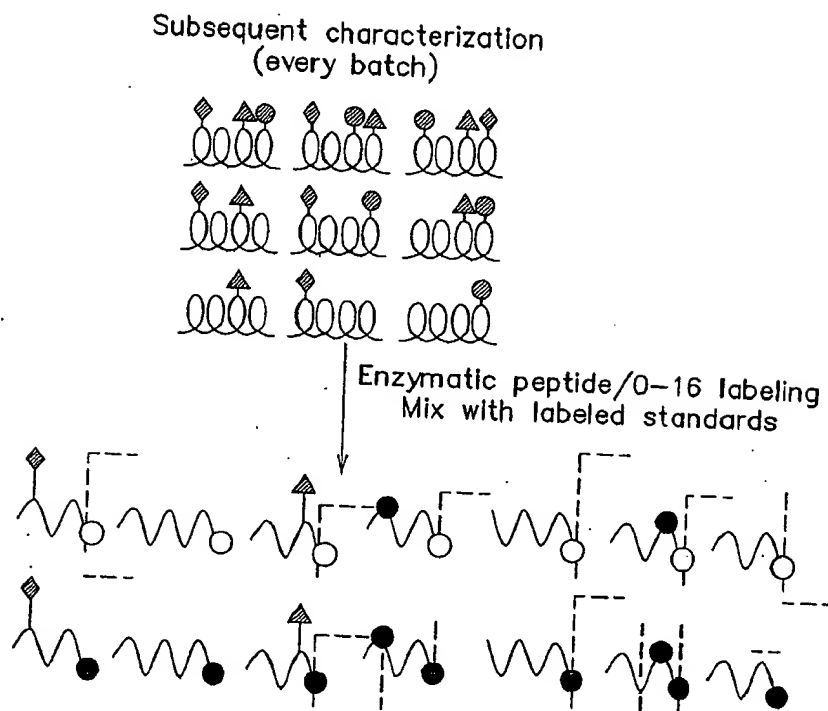


FIG. 24A

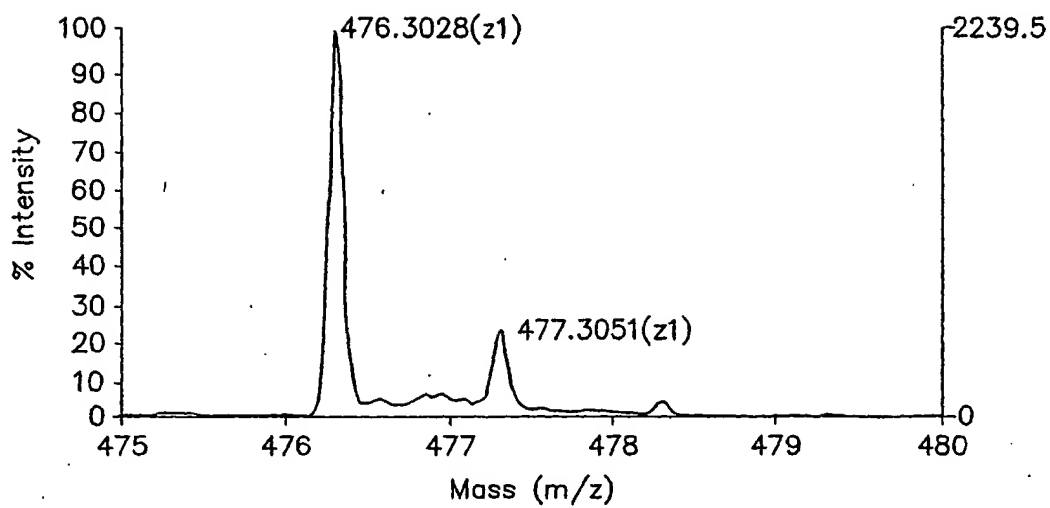
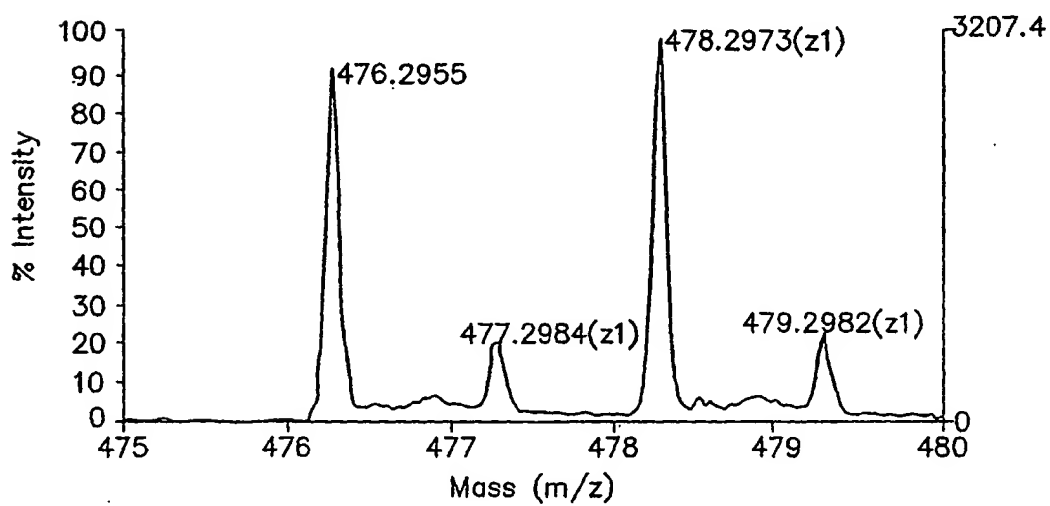
43/64



- Information from MS characterization:
- Quantitative glycan occupancy at each site
 - Quantity of each different glycan at each site
 - Preliminary ID of glycan structure at each site

FIG. 24B

44/64

**FIG. 25A****FIG. 25B**

45/64

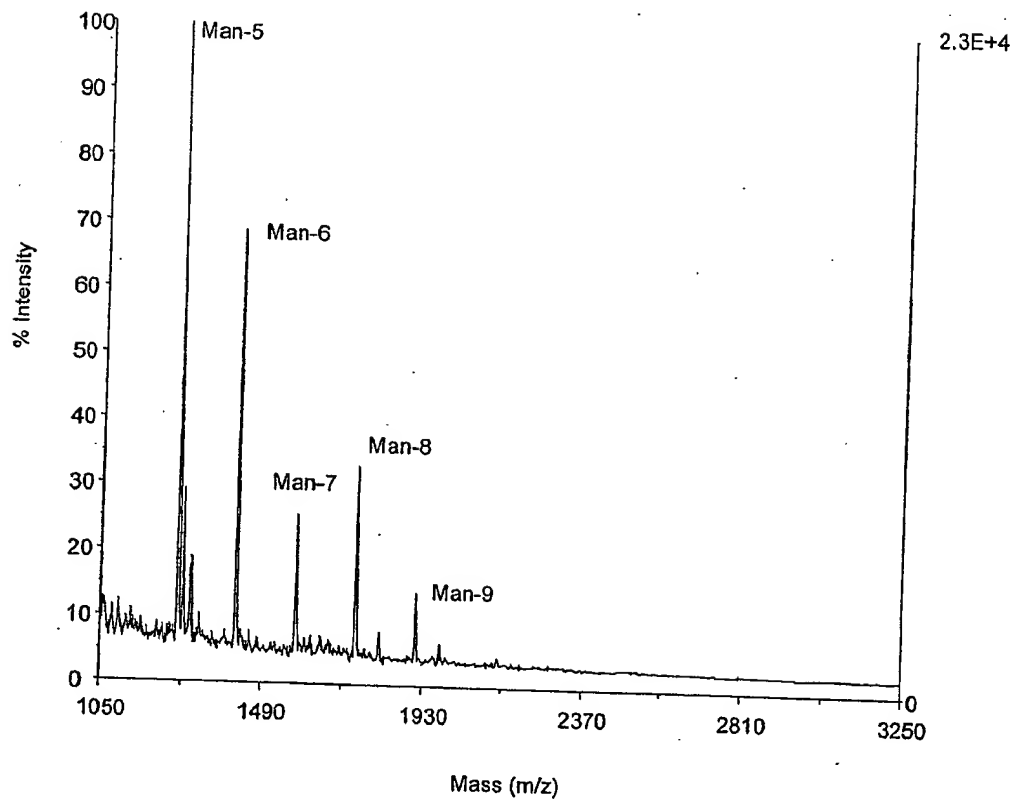


FIG. 26

46/64

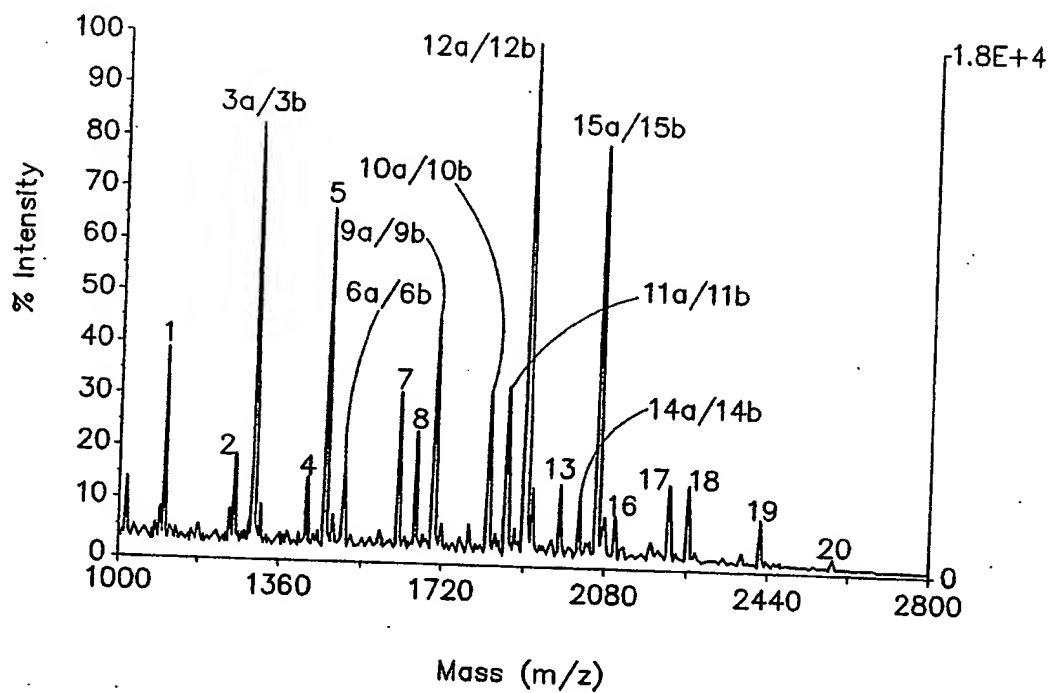


FIG. 27

47/64

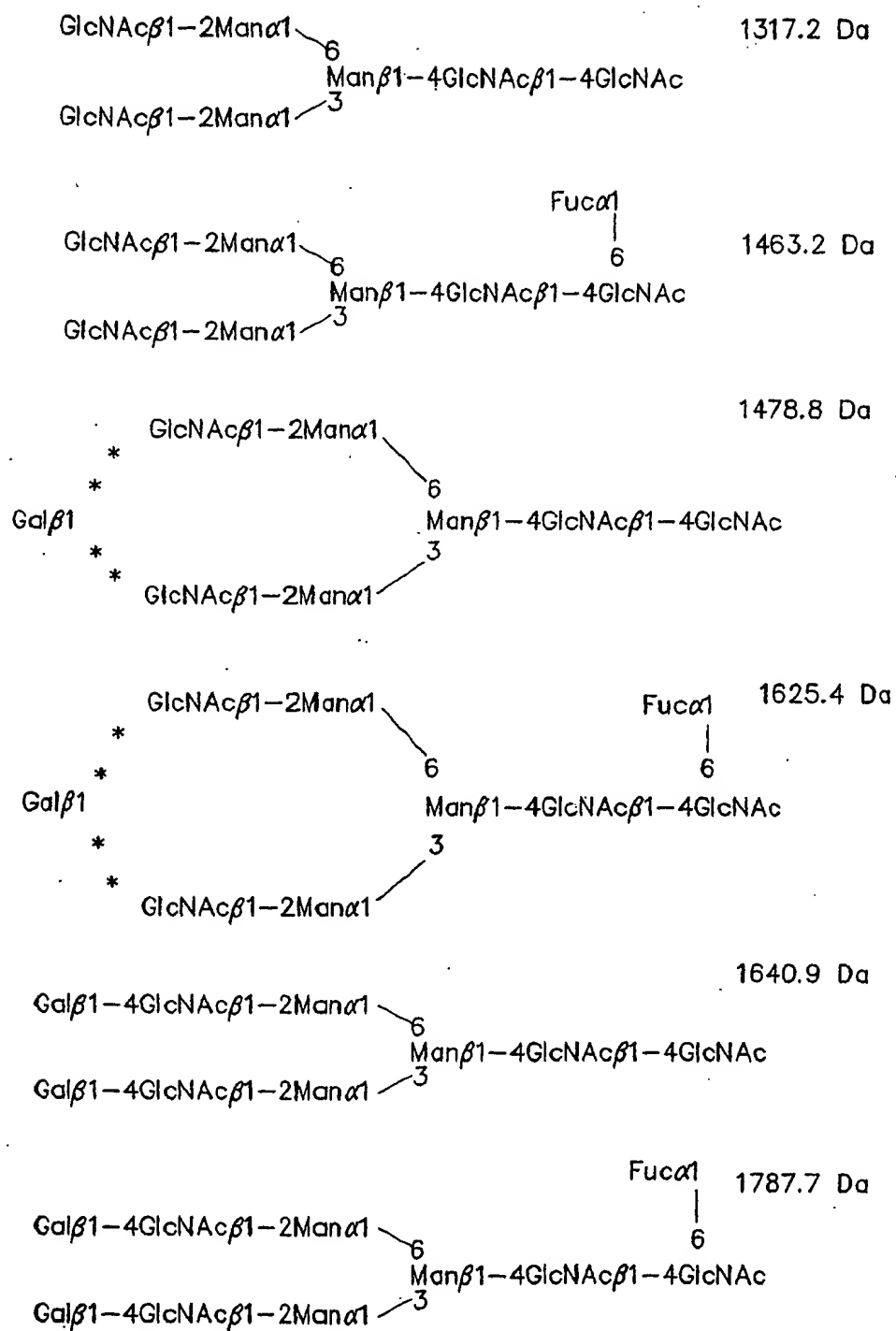


FIG. 28

48/64

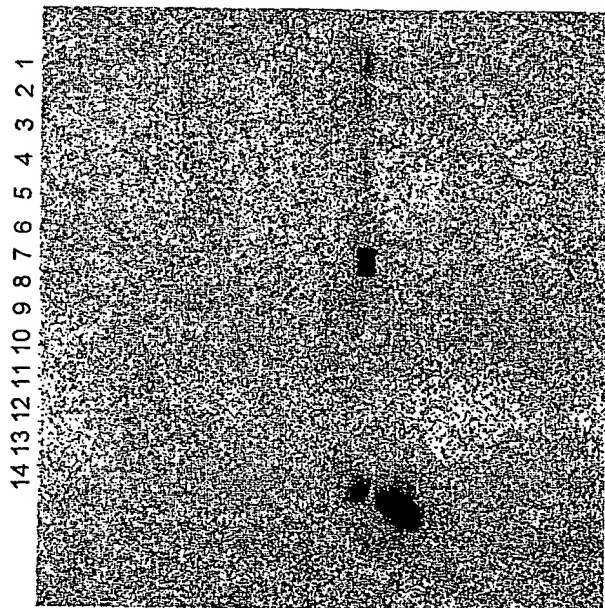


FIG. 29B

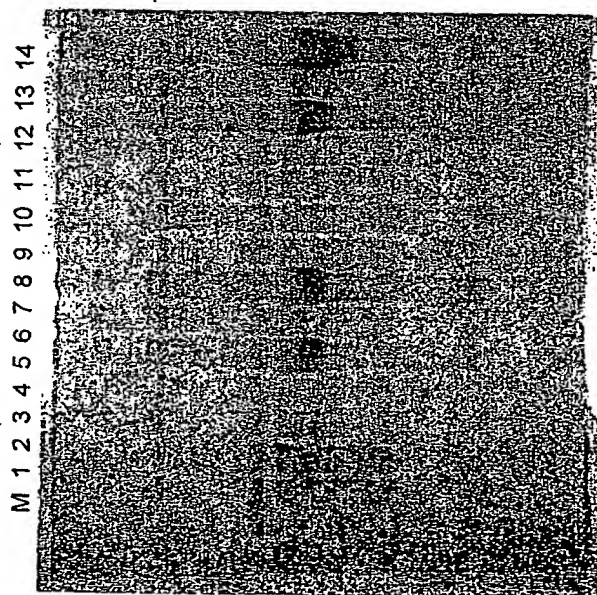


FIG. 29A

49/64

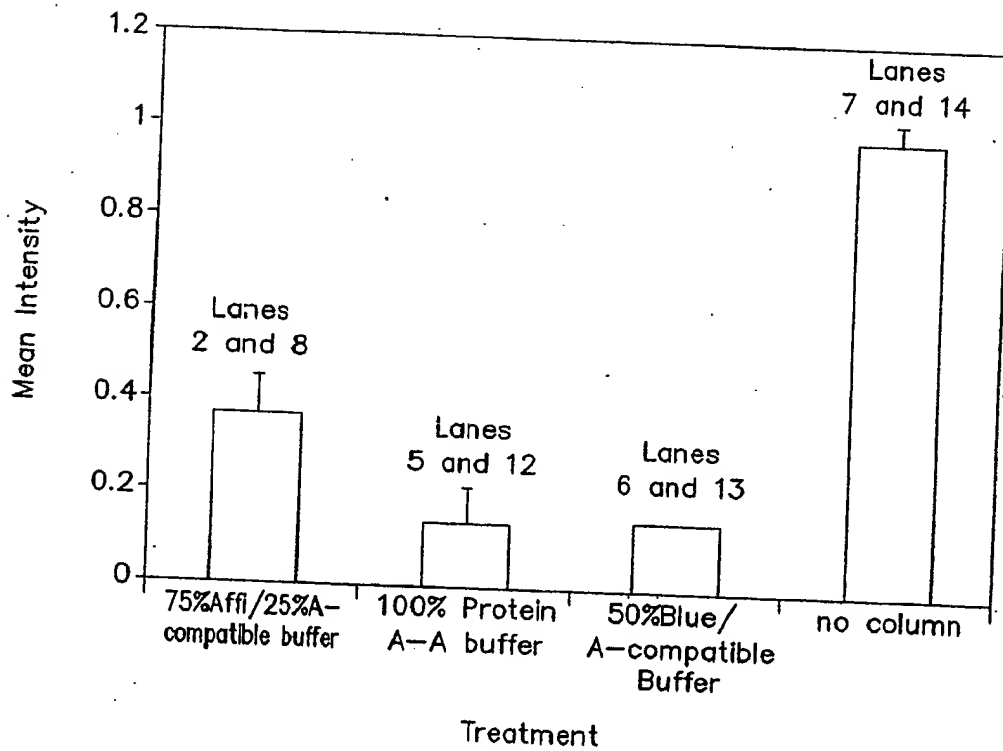


FIG. 29C

50/64

*FIG. 30*

51/64

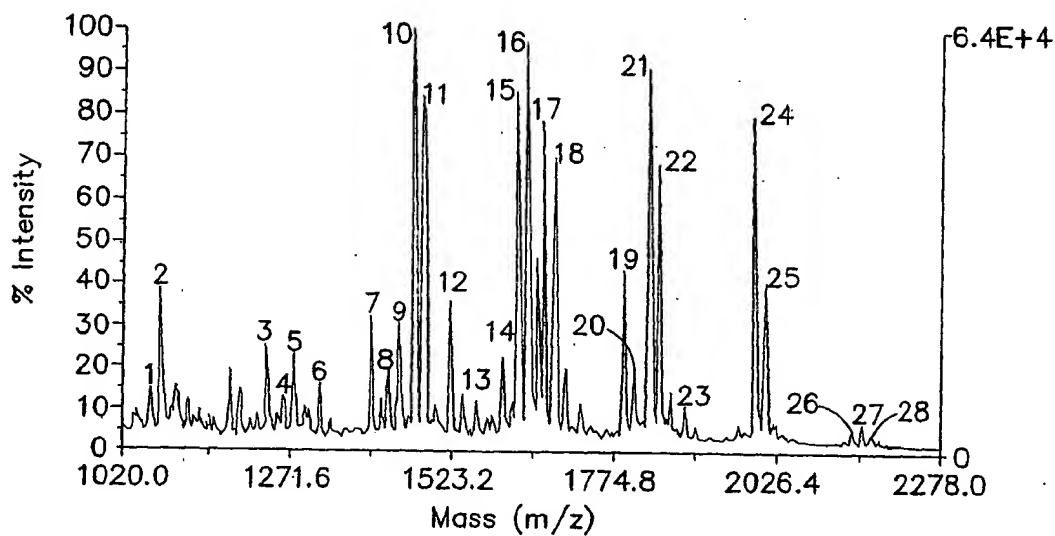


FIG. 31A

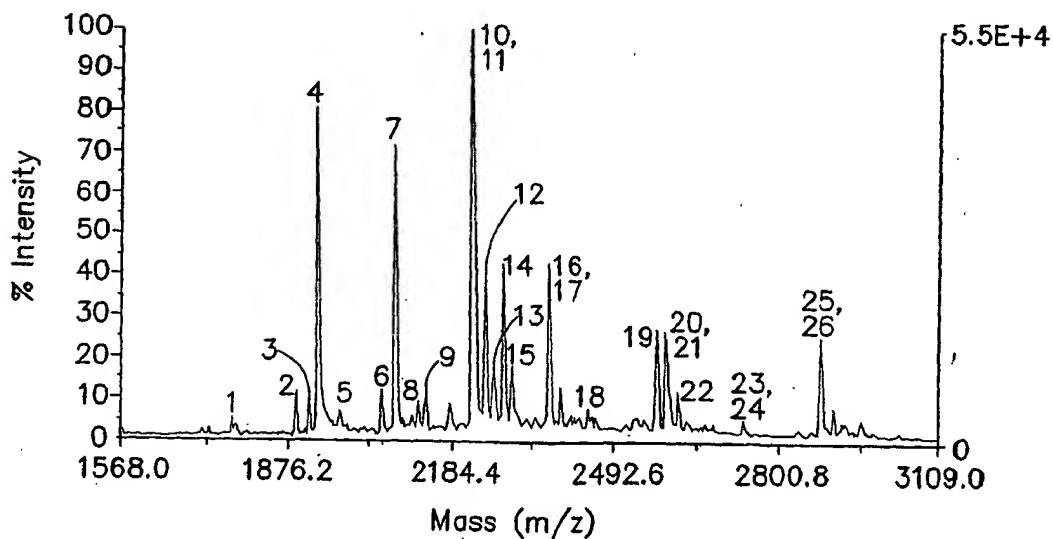


FIG. 31B

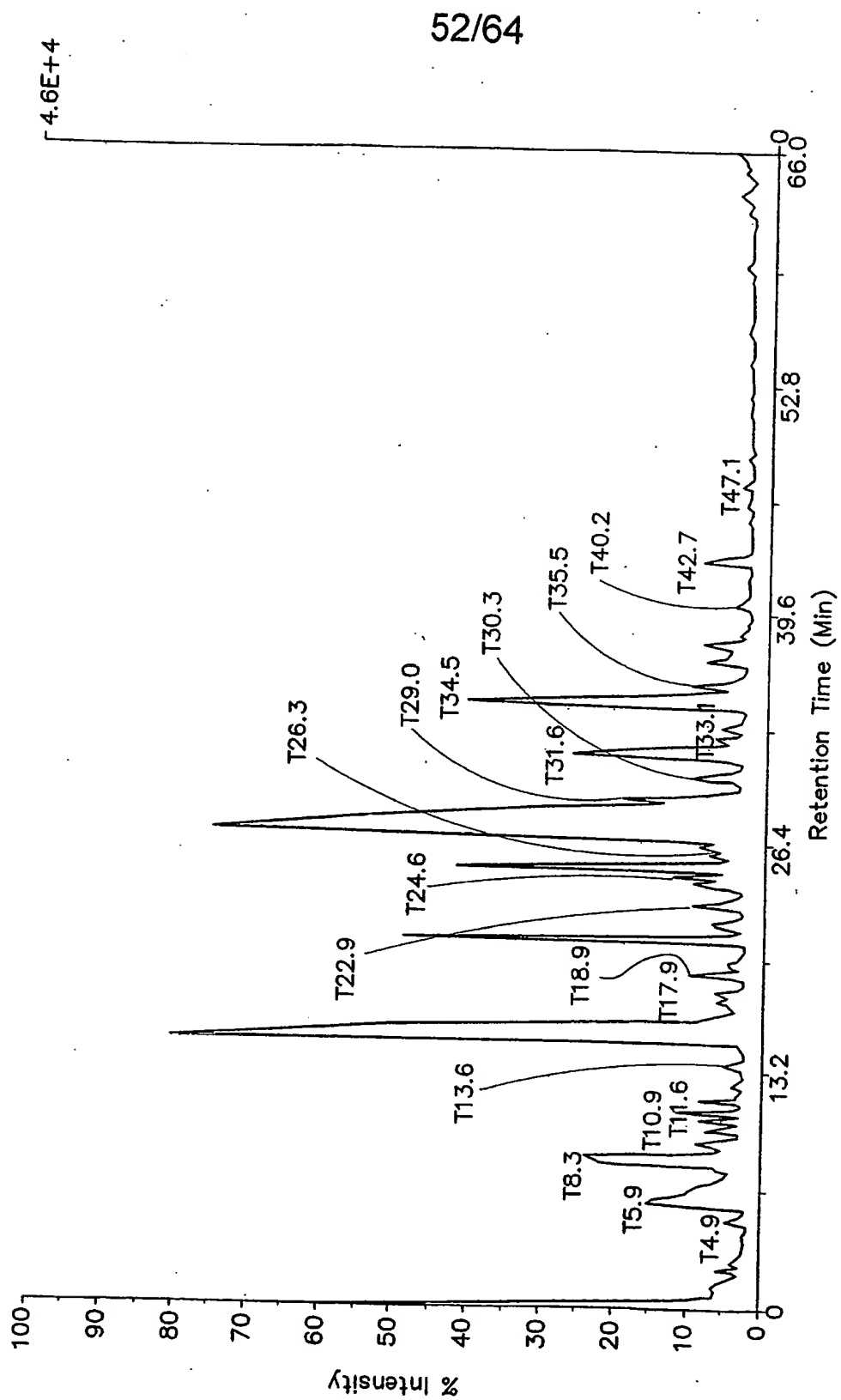


FIG. 32A

53/64

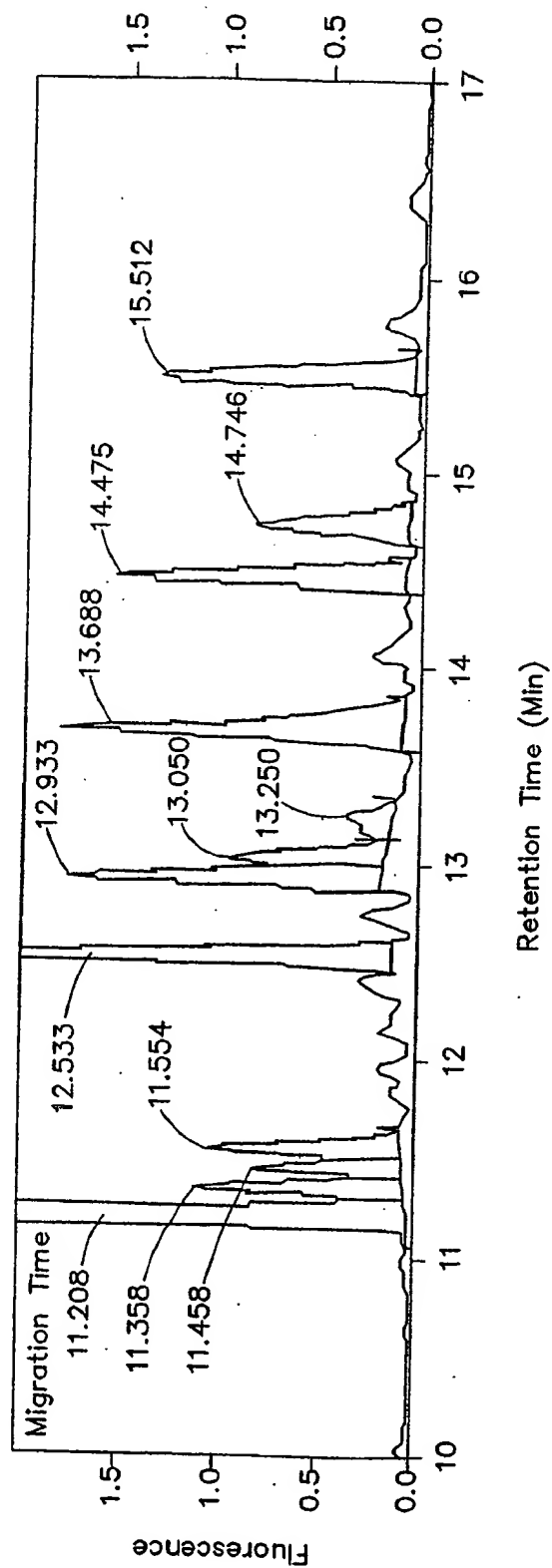


FIG. 32B

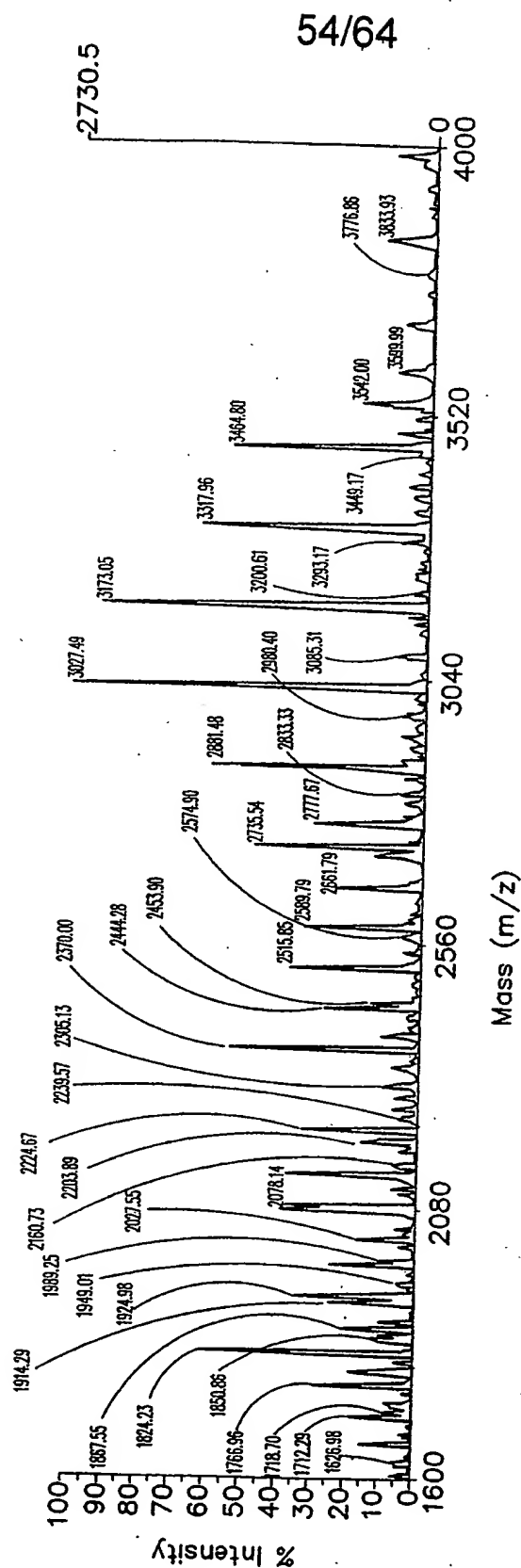


FIG. 33A

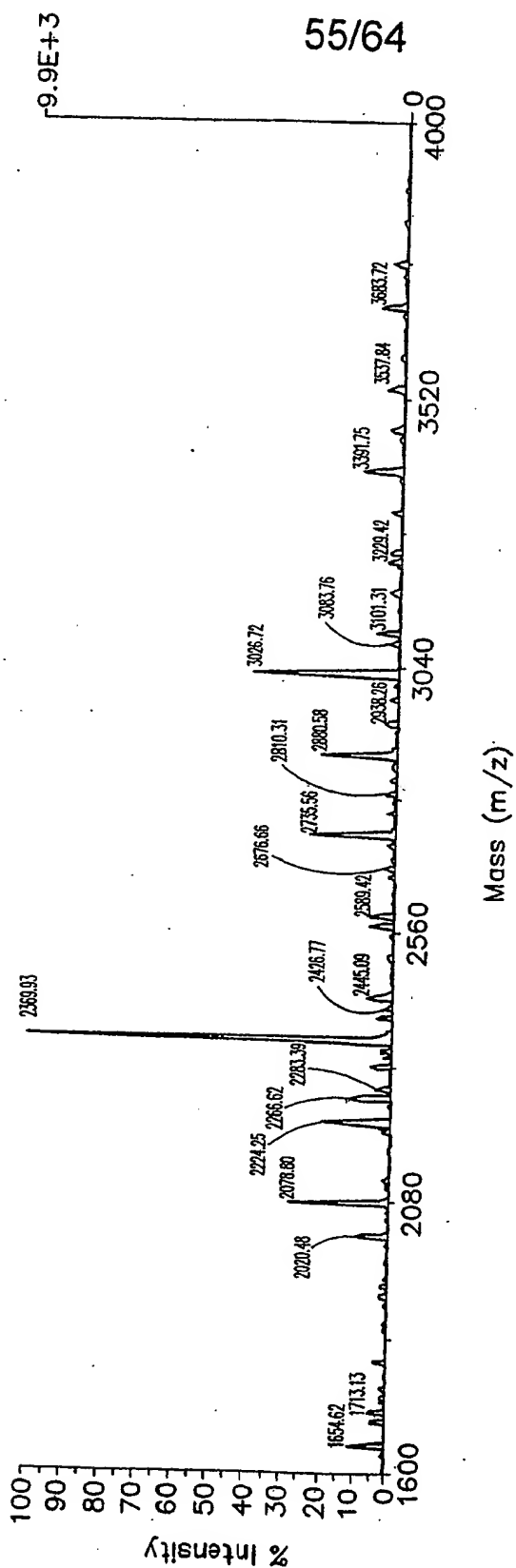


FIG. 33B

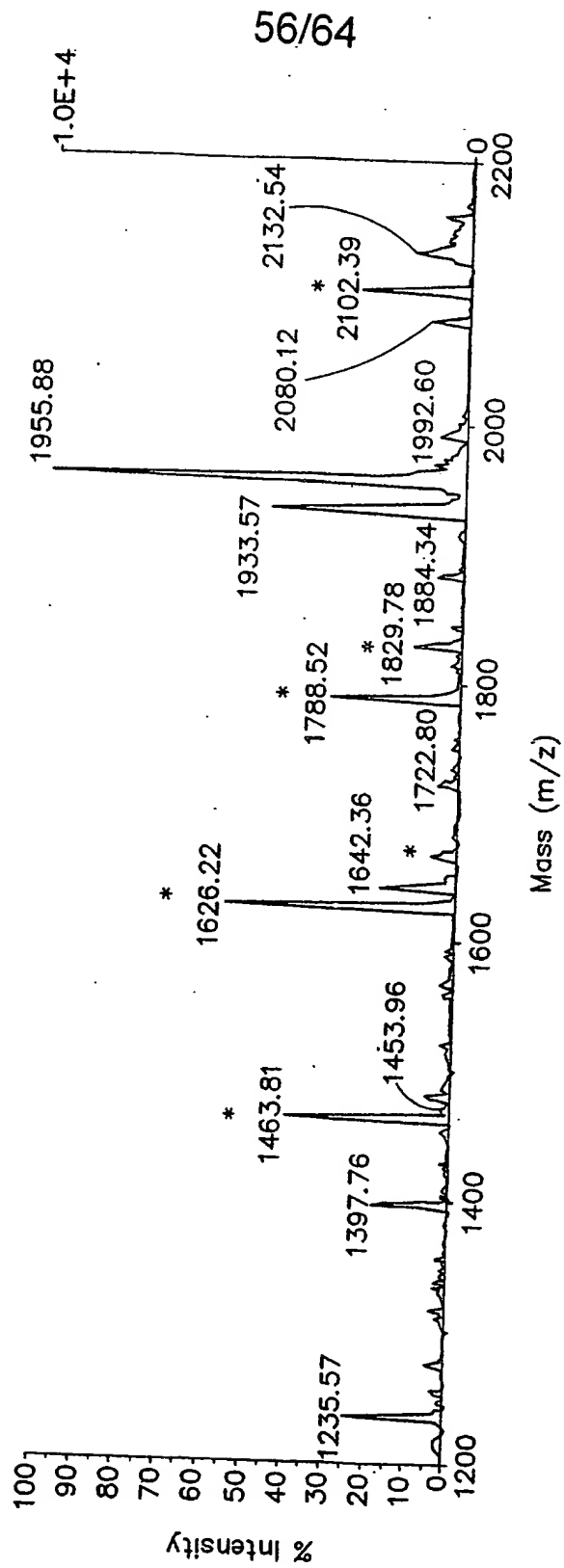


FIG. 34A

57/64

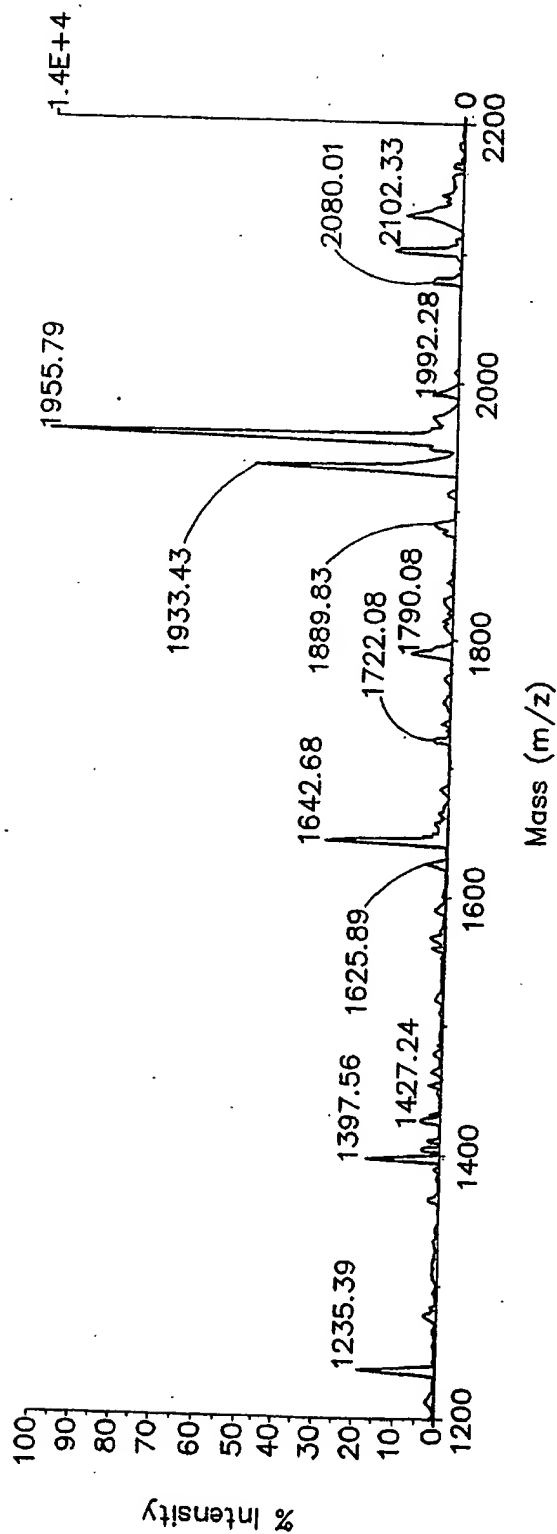


FIG. 34B

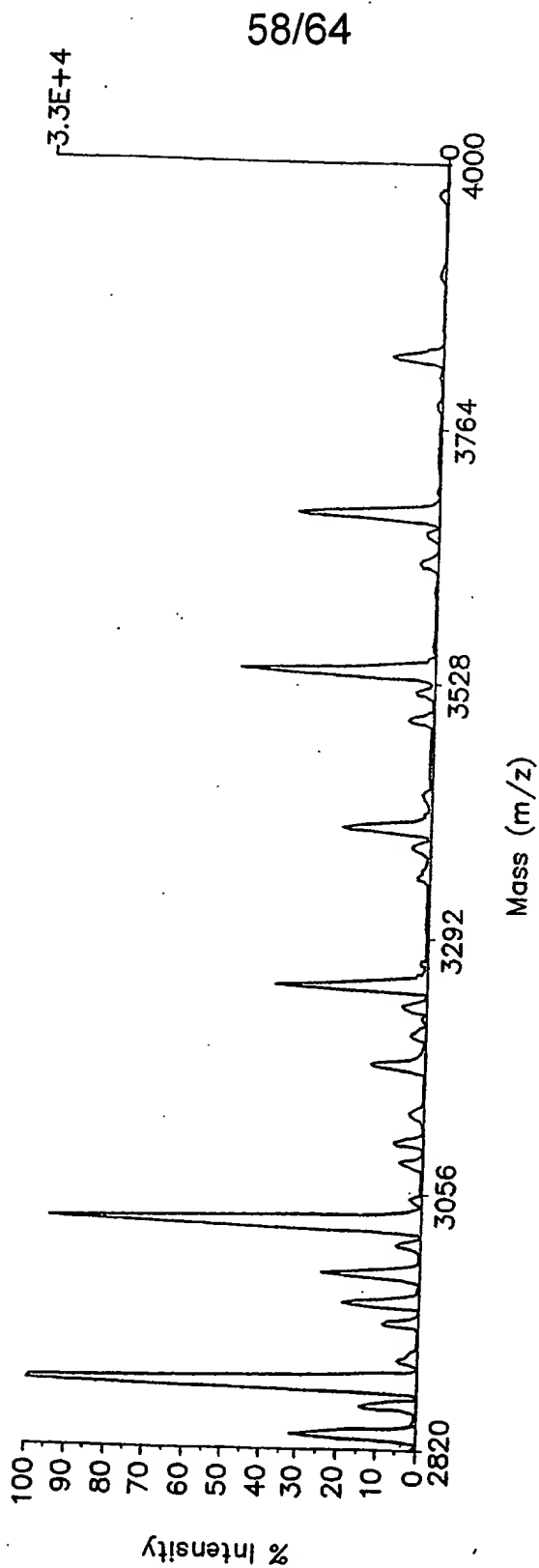


FIG. 35A

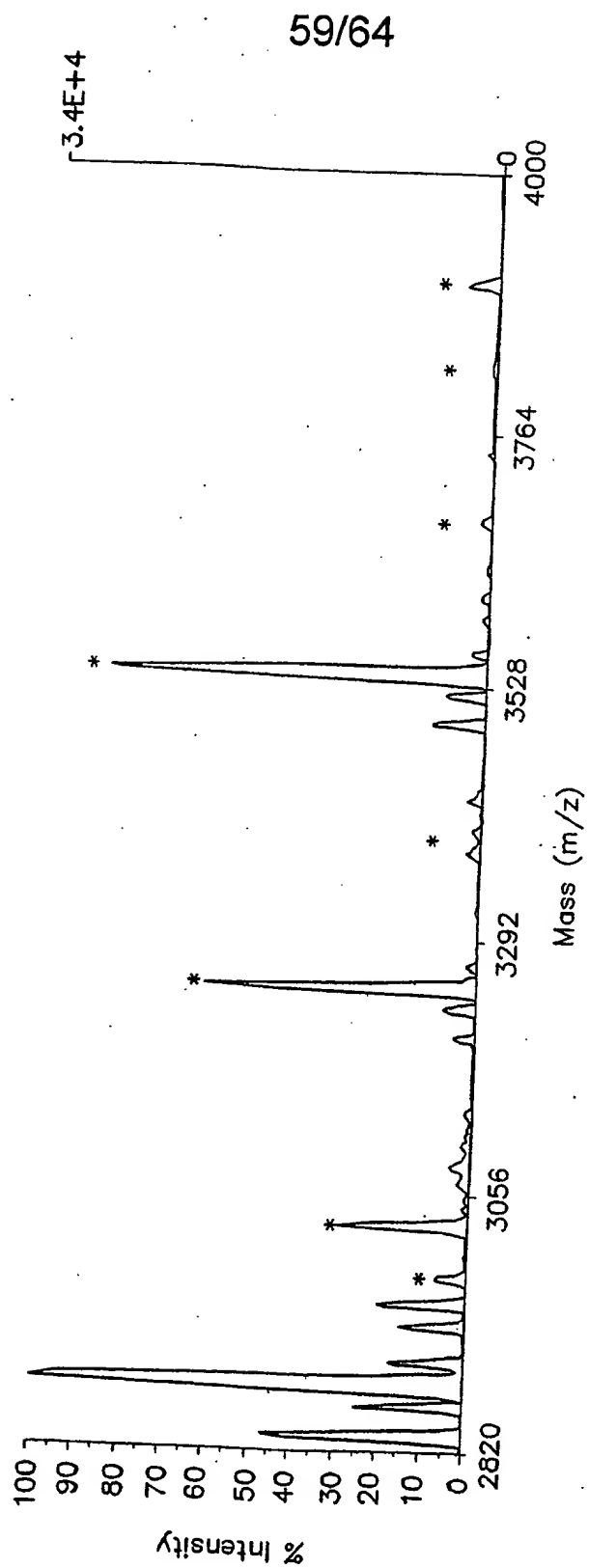


FIG. 35B

60/64

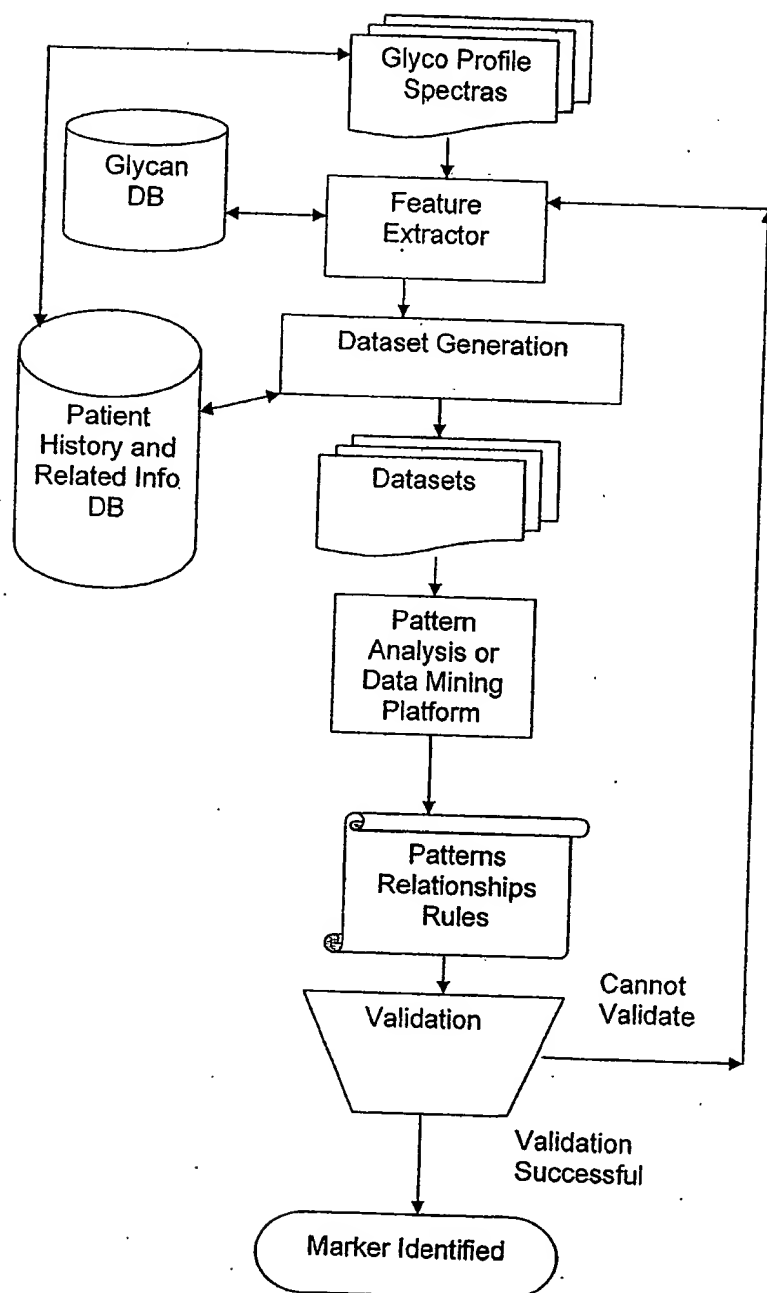


FIG. 36

61/64

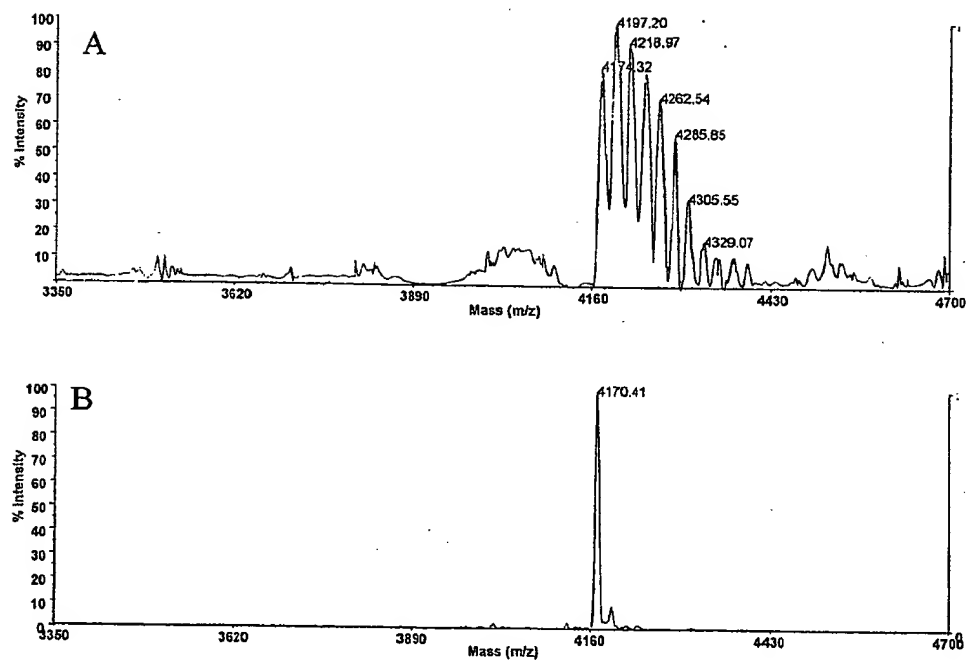


Fig. 37

62/64

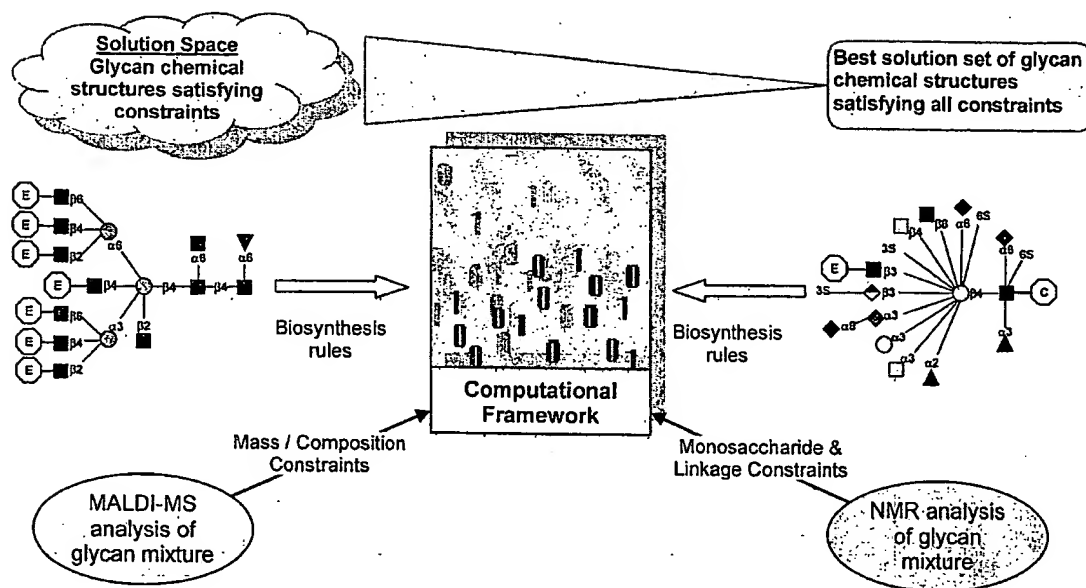
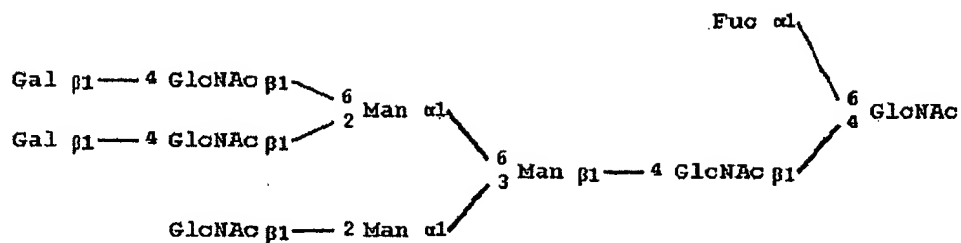


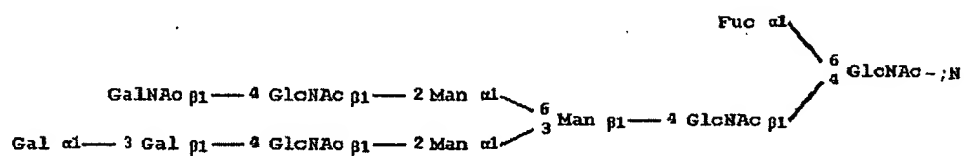
Fig. 38

63/64



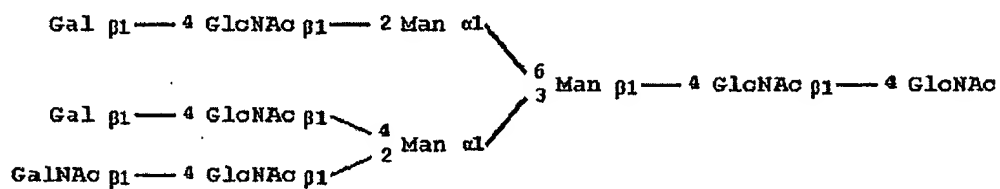
Mol Wt 1990; Man3Gal2Fuc1GlcNAc5 – 25%

A



Mol Wt 1990; Man3Gal2GlcNAc4GalNAc1Fuc1 – 50%

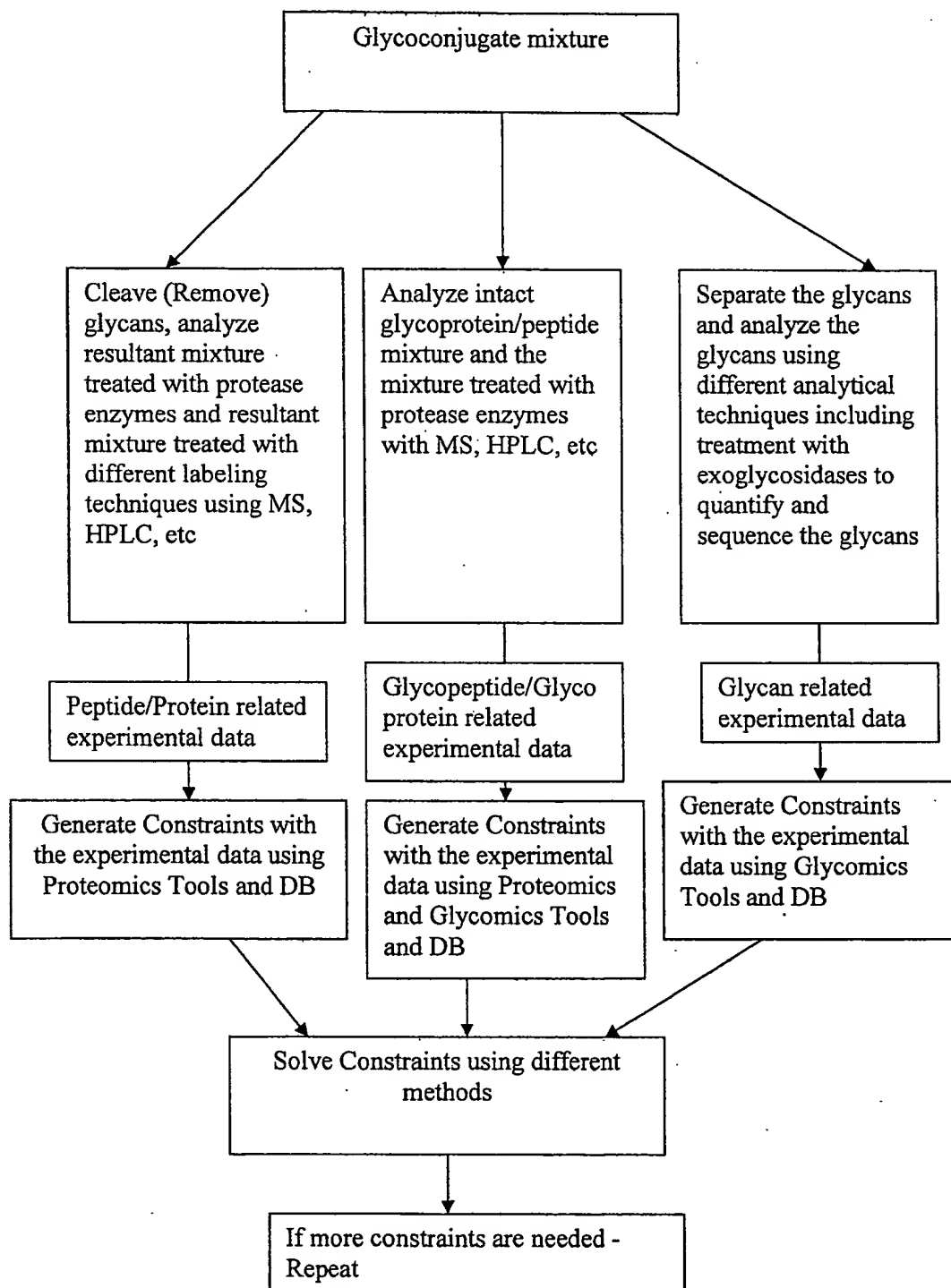
B



Mol Wt 2047; Man3Gal2GalNAc1GlcNAc5 – 25%

C

Fig. 39



SEQUENCE LISTING

<110> Massachusetts Institute of Technology et al.
<120> METHODS AND PRODUCTS RELATED TO THE IMPROVED ANALYSIS OF
CARBOHYDRATES
<130> M0656.70136W000
<150> US 11/244826
<151> 2005-10-06
<160> 8
<170> PatentIn version 3.3
<210> 1
<211> 17
<212> PRT
<213> Artificial sequence
<220>
<223> Synthetic peptide
<400> 1

Tyr Cys Asn Ile Ser Gln Lys Met Met Ser Arg Asn Leu Thr Lys Asp
1 5 10 15

Arg

<210> 2
<211> 4
<212> PRT
<213> Artificial sequence
<220>
<223> Synthetic peptide
<400> 2

Cys Asn Ile Ser
1

<210> 3
<211> 4
<212> PRT
<213> Artificial sequence
<220>
<223> Synthetic peptide
<400> 3

Arg Asn Leu Thr
1

<210> 4
<211> 4
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic peptide

<400> 4

Asn Leu Thr Lys
1

<210> 5
<211> 4
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic peptide

<400> 5

Met Met Ser Arg
1

<210> 6
<211> 7
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic peptide

<400> 6

Tyr Cys Asn Ile Ser Gln Lys
1 5

<210> 7
<211> 150
<212> PRT
<213> Bos taurus

<400> 7

Met Ala Leu Lys Ser Leu Val Leu Leu Ser Leu Leu Val Leu Val Leu
1 5 10 15

Leu Leu Val Arg Val Gln Pro Ser Leu Gly Lys Glu Thr Ala Ala Ala
20 25 30

Lys Phe Glu Arg Gln His Met Asp Ser Ser Thr Ser Ala Ala Ser Ser
35 40 45

Ser Asn Tyr Cys Asn Gln Met Met Lys Ser Arg Asn Leu Thr ^{64'}Asp
50 55 60

Arg Cys Lys Pro Val Asn Thr Phe Val His Glu Ser Leu Ala Asp Val
65 70 75 80

Gln Ala Val Cys Ser Gln Lys Asn Val Ala Cys Lys Asn Gly Gln Thr
85 90 95

Asn Cys Tyr Gln Ser Tyr Ser Thr Met Ser Ile Thr Asp Cys Arg Glu
100 105 110

Thr Gly Ser Ser Lys Tyr Pro Asn Cys Ala Tyr Lys Thr Thr Gln Ala
115 120 125

Asn Lys His Ile Ile Val Ala Cys Glu Gly Asn Pro Tyr Val Pro Val
130 135 140

His Phe Asp Ala Ser Val
145 150

<210> 8
<211> 4
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic peptide

<400> 8

Ser Asn Leu Thr
1